

PREDICTING BREAST CANCER TREATMENT OUTCOME

Related Applications

This application claims benefit of priority from U.S. Provisional Patent Application 60/504,087, filed September 19, 2003, and is a continuation in part of U.S. Patent Application
5 10/727,100, filed December 2, 2003. Both applications are hereby incorporated by reference in their entireties as if fully set forth.

Field of the Invention

The invention relates to the identification and use of gene expression profiles, or patterns, with clinical relevance to the treatment of breast cancer using tamoxifen (nolvadex) and other
10 “antiestrogen” agents against breast cancer, including other “selective estrogen receptor modulators” (“SERM”s), “selective estrogen receptor downregulators” (“SERD”s), and aromatase inhibitors (“AI”s). In particular, the invention provides the identities of gene sequences the expression of which are correlated with patient survival and breast cancer recurrence in women treated with tamoxifen or other “antiestrogen” agents against breast cancer. The gene expression
15 profiles, whether embodied in nucleic acid expression, protein expression, or other expression formats, may be used to select subjects afflicted with breast cancer who will likely respond positively to treatment with tamoxifen or another “antiestrogen” agent against breast cancer as well as those who will likely be non-responsive and thus candidates for other treatments. The invention also provides the identities of three sets of sequences from three genes with expression patterns that
20 are strongly predictive of responsiveness to tamoxifen and other “antiestrogen” agents against breast cancer.

Background of the Invention

Breast cancer is by far the most common cancer among women. Each year, more than 180,000 and 1 million women in the U.S. and worldwide, respectively, are diagnosed with breast
25 cancer. Breast cancer is the leading cause of death for women between ages 50-55, and is the most

common non-preventable malignancy in women in the Western Hemisphere. An estimated 2,167,000 women in the United States are currently living with the disease (National Cancer Institute, Surveillance Epidemiology and End Results (NCI SEER) program, *Cancer Statistics Review (CSR)*, www-seer.ims.nci.nih.gov/Publications/CSR1973 (1998)). Based on cancer rates from 1995 through 1997, a report from the National Cancer Institute (NCI) estimates that about 1 in 8 women in the United States (approximately 12.8 percent) will develop breast cancer during her lifetime (NCI's Surveillance, Epidemiology, and End Results Program (SEER) publication *SEER Cancer Statistics Review 1973-1997*). Breast cancer is the second most common form of cancer, after skin cancer, among women in the United States. An estimated 250,100 new cases of breast cancer are expected to be diagnosed in the United States in 2001. Of these, 192,200 new cases of more advanced (invasive) breast cancer are expected to occur among women (an increase of 5% over last year), 46,400 new cases of early stage (*in situ*) breast cancer are expected to occur among women (up 9% from last year), and about 1,500 new cases of breast cancer are expected to be diagnosed in men (Cancer Facts & Figures 2001 American Cancer Society). An estimated 40,600 deaths (40,300 women, 400 men) from breast cancer are expected in 2001. Breast cancer ranks second only to lung cancer among causes of cancer deaths in women. Nearly 86% of women who are diagnosed with breast cancer are likely to still be alive five years later, though 24% of them will die of breast cancer after 10 years, and nearly half (47%) will die of breast cancer after 20 years.

Every woman is at risk for breast cancer. Over 70 percent of breast cancers occur in women who have no identifiable risk factors other than age (U.S. General Accounting Office. Breast Cancer, 1971-1991: Prevention, Treatment and Research. GAO/PEMD-92-12; 1991). Only 5 to 10% of breast cancers are linked to a family history of breast cancer (Henderson IC, Breast Cancer. In: Murphy GP, Lawrence WL, Lenhard RE (eds). *Clinical Oncology*. Atlanta, GA: American Cancer Society; 1995:198-219).

Each breast has 15 to 20 sections called lobes. Within each lobe are many smaller lobules. Lobules end in dozens of tiny bulbs that can produce milk. The lobes, lobules, and bulbs are all linked by thin tubes called ducts. These ducts lead to the nipple in the center of a dark area of skin called the areola. Fat surrounds the lobules and ducts. There are no muscles in the breast, but muscles lie under each breast and cover the ribs. Each breast also contains blood vessels and lymph

vessels. The lymph vessels carry colorless fluid called lymph, and lead to the lymph nodes. Clusters of lymph nodes are found near the breast in the axilla (under the arm), above the collarbone, and in the chest.

Breast tumors can be either benign or malignant. Benign tumors are not cancerous, they do not spread to other parts of the body, and are not a threat to life. They can usually be removed, and in most cases, do not come back. Malignant tumors are cancerous, and can invade and damage nearby tissues and organs. Malignant tumor cells may metastasize, entering the bloodstream or lymphatic system. When breast cancer cells metastasize outside the breast, they are often found in the lymph nodes under the arm (axillary lymph nodes). If the cancer has reached these nodes, it means that cancer cells may have spread to other lymph nodes or other organs, such as bones, liver, or lungs.

Major and intensive research has been focused on early detection, treatment and prevention. This has included an emphasis on determining the presence of precancerous or cancerous ductal epithelial cells. These cells are analyzed, for example, for cell morphology, for protein markers, for nucleic acid markers, for chromosomal abnormalities, for biochemical markers, and for other characteristic changes that would signal the presence of cancerous or precancerous cells. This has led to various molecular alterations that have been reported in breast cancer, few of which have been well characterized in human clinical breast specimens. Molecular alterations include presence/absence of estrogen and progesterone steroid receptors, HER-2 expression/amplification (Mark HF, et al. HER-2/neu gene amplification in stages I-IV breast cancer detected by fluorescent in situ hybridization. *Genet Med*; 1(3):98-103 1999), Ki-67 (an antigen that is present in all stages of the cell cycle except G0 and used as a marker for tumor cell proliferation, and prognostic markers (including oncogenes, tumor suppressor genes, and angiogenesis markers) like p53, p27, Cathepsin D, pS2, multi-drug resistance (MDR) gene, and CD31.

Tamoxifen is the antiestrogen agent most frequently prescribed in women with both early stage and metastatic hormone receptor-positive breast cancer (for reviews, see Clarke, R. et al. "Antiestrogen resistance in breast cancer and the role of estrogen receptor signaling." *Oncogene* 22, 7316-39 (2003) and Jordan, C. "Historical perspective on hormonal therapy of advanced breast Cancer." *Clin. Ther.* 24 Suppl A, A3-16 (2002)). In the adjuvant setting, tamoxifen therapy results

in a 40-50% reduction in the annual risk of recurrence, leading to a 5.6% improvement in 10 year survival in lymph node negative patients, and a corresponding 10.9% improvement in node-positive patients (Group, E.B.C.T.C. Tamoxifen for early breast cancer. Cochrane Database Syst Rev, CD000486 (2001)). Tamoxifen is thought to act primarily as a competitive inhibitor of estrogen binding to estrogen receptor (ER). The absolute levels of ER expression, as well as that of the progesterone receptor (PR, an indicator of a functional ER pathway), are currently the best predictors of tamoxifen response in the clinical setting (Group, (2001) and Bardou, V.J. et al. "Progesterone receptor status significantly improves outcome prediction over estrogen receptor status alone for adjuvant endocrine therapy in two large breast cancer databases." J Clin Oncol 21, 1973-9 (2003)).

However, 25% of ER+/PR+ tumors, 66% of ER+/PR- cases and 55% of ER-/PR+ cases fail to respond, or develop early resistance to tamoxifen, through mechanisms that remain largely unclear (see Clarke et al.; Nicholson, R.I. et al. "The biology of antihormone failure in breast cancer." Breast Cancer Res Treat 80 Suppl 1, S29-34; discussion S35 (2003) and Osborne, C.K. et al. "Growth factor receptor cross-talk with estrogen receptor as a mechanism for tamoxifen resistance in breast cancer." Breast 12, 362-7 (2003)). Currently, no reliable means exist to allow the identification of these non-responders. In these patients, the use of alternative hormonal therapies, such as the aromatase inhibitors letrozole and anastrozole (Ellis, M.J. et al. "Letrozole is more effective neoadjuvant endocrine therapy than tamoxifen for ErbB-1- and/or ErbB-2-positive, estrogen receptorpositive primary breast cancer: evidence from a phase III randomized trial." J Clin Oncol 19, 3808-16 (2001); Buzdar, A.U. "Anastrozole: a new addition to the armamentarium against advanced breast cancer." Am J Clin Oncol 21, 161-6 (1998); and Goss, P.E. et al. "A randomized trial of letrozole in postmenopausal women after five years of tamoxifen therapy for early-stage breast cancer." N Engl J Med 349, 1793-802 (2003)); chemotherapeutic agents, or inhibitors of other signaling pathways, such as trastuzumab and gefitinib might offer the possibility of improving clinical outcome. Therefore, the ability to accurately predict tamoxifen treatment outcome should significantly advance the management of early stage breast cancer by identifying patients who are unlikely to benefit from TAM so that additional or alternative therapies may be sought.

Citation of documents herein is not intended as an admission that any is pertinent prior art. All statements as to the date or representation as to the contents of documents is based on the information available to the applicant and does not constitute any admission as to the correctness of the dates or contents of the documents.

5 Summary of the Invention

 The present invention relates to the identification and use of gene expression patterns (or profiles or “signatures”) and the expression levels of individual gene sequences which are clinically relevant to breast cancer. In particular, the identities of genes that are correlated with patient survival and breast cancer recurrence (e.g. metastasis of the breast cancer) are provided. The gene
10 expression profiles, whether embodied in nucleic acid expression, protein expression, or other expression formats, may be used to predict survival of subjects afflicted with breast cancer and the likelihood of breast cancer recurrence, including cancer metastasis.

 The invention thus provides for the identification and use of gene expression patterns (or profiles or “signatures”) and the expression levels of individual gene sequences which correlate
15 with (and thus are able to discriminate between) patients with good or poor survival outcomes. In one embodiment, the invention provides patterns that are able to distinguish patients with estrogen receptor (α isoform) positive (ER+) breast tumors into those with that are responsive, or likely to be responsive, to treatment with tamoxifen (TAM) or another “antiestrogen” agent against breast cancer (such as a “selective estrogen receptor modulator” (“SERM”), “selective estrogen receptor
20 downregulator” (“SERD”), or aromatase inhibitor (“AI”)) and those that are non-responsive, or likely to be non-responsive, to such treatment. In an alternative embodiment, the invention may be applied to patients with breast tumors that do not display detectable levels of ER expression (so called “ER-“ subjects) but where the patient will nonetheless benefit from application of the invention due to the presence of some low level ER expression. Responsiveness may be viewed in
25 terms of better survival outcomes over time. These patterns are thus able to distinguish patients with ER+ breast tumors into at least two subtypes.

 In a first aspect, the present invention provides a non-subjective means for the identification of patients with breast cancer (ER+ or ER-) as likely to have a good or poor survival outcome

following treatment with TAM or another “antiestrogen” agent against breast cancer by assaying for the expression patterns disclosed herein. Thus where subjective interpretation may have been previously used to determine the prognosis and/or treatment of breast cancer patients, the present invention provides objective gene expression patterns, which may used alone or in combination with subjective criteria to provide a more accurate assessment of ER+ or ER- breast cancer patient outcomes or expected outcomes, including survival and the recurrence of cancer, following treatment with TAM or another “antiestrogen” agent against breast cancer. The expression patterns of the invention thus provide a means to determine ER+ or ER- breast cancer prognosis. Furthermore, the expression patterns can also be used as a means to assay small, node negative tumors that are not readily assayed by other means.

The gene expression patterns comprise one or more than one gene capable of discriminating between breast cancer outcomes with significant accuracy. The gene sequence(s) are identified as correlated with ER+ breast cancer outcomes such that the levels of their expression are relevant to a determination of the preferred treatment protocols for a patient, whether ER+ or ER-. Thus in one embodiment, the invention provides a method to determine the outcome of a subject afflicted with breast cancer by assaying a cell containing sample from said subject for expression of one or more than one gene disclosed herein as correlated with breast cancer outcomes following treatment with TAM or another “antiestrogen” agent against breast cancer.

The ability to correlate gene expression with breast cancer outcome and responsiveness to TAM is particularly advantageous in light of the possibility that up to 40% of ER+ subjects that undergo TAM treatment are non-responders. Therefore, the ability to identify, with confidence, these non-responders at an early time point permits the consideration and/or application of alternative therapies (such as a different “antiestrogen” agent against breast cancer or other anti-breast cancer treatments) to the non-responders. Stated differently, the ability to identify TAM non-responder subjects permits medical personnel to consider and/or utilize alternative therapies for the treatment of the subjects before time is spent on ineffective TAM therapy. Time spent on an ineffective therapy often permits further cancer growth, and the likelihood of success with alternative therapies diminishes over time given such growth. Therefore, the invention also

provides methods to improve the survival outcome of non-responders by use of the methods disclosed herein to identify non-responders for treatment with alternative therapies.

Gene expression patterns of the invention are identified as described below. Generally, a large sampling of the gene expression profile of a sample is obtained through quantifying the expression levels of mRNA corresponding to many genes. This profile is then analyzed to identify genes, the expression of which are positively, or negatively, correlated, with ER+ breast cancer outcome upon treatment with TAM or another “antiestrogen” agent against breast cancer. An expression profile of a subset of human genes may then be identified by the methods of the present invention as correlated with a particular outcome. The use of multiple samples increases the confidence which a gene may be believed to be correlated with a particular survival outcome. Without sufficient confidence, it remains unpredictable whether expression of a particular gene is actually correlated with an outcome and also unpredictable whether expression of a particular gene may be successfully used to identify the outcome for a breast cancer patient. While the invention may be practiced based on the identities of the gene sequences disclosed herein or the actual sequences used independent of identification, the invention may also be practiced with any other sequences the expression of which is correlated with the expression of sequences disclosed herein. Such additional sequences may be identified by any means known in the art, including the methods disclosed herein.

A profile of genes that are highly correlated with one outcome relative to another may be used to assay an sample from a subject afflicted with breast cancer to predict the likely responsiveness (or lack thereof) to TAM or another “antiestrogen” agent against breast cancer in the subject from whom the sample was obtained. Such an assay may be used as part of a method to determine the therapeutic treatment for said subject based upon the breast cancer outcome identified.

As discussed below, the correlated genes may be used singly with significant accuracy or in combination to increase the ability to accurately correlating a molecular expression phenotype with a breast cancer outcome. This correlation is a way to molecularly provide for the determination of survival outcomes as disclosed herein. Additional uses of the correlated gene(s) are in the

classification of cells and tissues; determination of diagnosis and/or prognosis; and determination and/or alteration of therapy.

The ability to discriminate is conferred by the identification of expression of the individual genes as relevant and not by the form of the assay used to determine the actual level of expression.

5 An assay may utilize any identifying feature of an identified individual gene as disclosed herein as long as the assay reflects, quantitatively or qualitatively, expression of the gene in the “transcriptome” (the transcribed fraction of genes in a genome) or the “proteome” (the translated fraction of expressed genes in a genome). Additional assays include those based on the detection of polypeptide fragments of the relevant member or members of the proteome. Identifying features
10 include, but are not limited to, unique nucleic acid sequences used to encode (DNA), or express (RNA), said gene or epitopes specific to, or activities of, a protein encoded by said gene. All that is required are the gene sequence(s) necessary to discriminate between breast cancer outcomes and an appropriate cell containing sample for use in an expression assay.

In another embodiment, the invention provides for the identification of the gene expression
15 patterns by analyzing global, or near global, gene expression from single cells or homogenous cell populations which have been dissected away from, or otherwise isolated or purified from, contaminating cells beyond that possible by a simple biopsy. Because the expression of numerous genes fluctuate between cells from different patients as well as between cells from the same patient sample, multiple data from expression of individual genes and gene expression patterns are used as
20 reference data to generate models which in turn permit the identification of individual gene(s), the expression of which are most highly correlated with particular breast cancer outcomes.

In additional embodiments, the invention provides physical and methodological means for detecting the expression of gene(s) identified by the models generated by individual expression patterns. These means may be directed to assaying one or more aspects of the DNA template(s)
25 underlying the expression of the gene(s), of the RNA used as an intermediate to express the gene(s), or of the proteinaceous product expressed by the gene(s).

In further embodiments, the gene(s) identified by a model as capable of discriminating between breast cancer outcomes may be used to identify the cellular state of an unknown sample of cell(s) from the breast. Preferably, the sample is isolated via non-invasive means. The expression

of said gene(s) in said unknown sample may be determined and compared to the expression of said gene(s) in reference data of gene expression patterns correlated with breast cancer outcomes.

Optionally, the comparison to reference samples may be by comparison to the model(s) constructed based on the reference samples.

5 One advantage provided by the present invention is that contaminating, non-breast cells (such as infiltrating lymphocytes or other immune system cells) are not present to possibly affect the genes identified or the subsequent analysis of gene expression to identify the survival outcomes of patients with breast cancer. Such contamination is present where a biopsy is used to generate gene expression profiles. However, and as noted herein, the invention includes the identity of genes
10 that may be used with significant accuracy even in the presence of contaminating cells.

 In a second aspect, the invention provides a non-subjective means based on the expression of three genes, or combinations thereof, for the identification of patients with breast cancer as likely to have a good or poor survival outcome following treatment with TAM or another “antiestrogen” agent against breast cancer. These three genes are members of the expression patterns disclosed
15 herein which have been found to be strongly predictive of clinical outcome following TAM treatment of ER+ breast cancer.

 The present invention thus provides gene sequences identified as differentially expressed in ER+ breast cancer in correlation to TAM responsiveness. The sequences of two of the genes display increased expression in ER+ breast cells that respond to TAM treatment (and thus lack of
20 increased expression in nonresponsive cases). The sequences of the third gene display decreased expression in ER+ breast cells that respond to TAM treatment (and thus lack of decreased expression in nonresponsive cases).

 The first set of sequences found to be more highly expressed in TAM responsive, ER+ breast cells are those of interleukin 17 receptor B (IL17RB), which has been mapped to human
25 chromosome 3 at 3p21.1. IL17RB is also referred to as interleukin 17B receptor (IL17BR) and sequences corresponding to it, and thus may be used in the practice of the instant invention, are identified by UniGene Cluster Hs.5470.

 The second set of sequences found to be more highly expressed in TAM responsive, ER+ breast cells are those of the calcium channel, voltage-dependent, L type, alpha 1D subunit

(CACNA1D), which has been mapped to human chromosome 3 at 3p14.3. Sequences corresponding to CACNA1D, and thus may be used in the practice of the instant invention, are identified by UniGene Cluster Hs.399966.

The set of sequences found to be expressed at lower levels in TAM responsive, ER+ breast cells are those of homeobox B13 (HOXB13), which has been mapped to human chromosome 17 at 17q21.2. Sequences corresponding to HOXB13, and thus may be used in the practice of the instant invention, are identified by UniGene Cluster Hs.66731.

While the invention may be practiced based on the identities of these three gene sequences or the actual sequences used independent of the assigned identity, the invention may also be practiced with any other sequence the expression of which is correlated with the expression of these disclosed sequences. Such additional sequences may be identified by any means known in the art, including the methods disclosed herein.

The identified sequences may thus be used in methods of determining the responsiveness, or non-responsiveness, of a subject's ER+ or ER- breast cancer to TAM treatment, or treatment with another "antiestrogen" agent against breast cancer, via analysis of breast cells in a tissue or cell containing sample from a subject. As non-limiting examples, the lack of increased expression of IL17BR and CACNA1D sequences and/or the lack of decreased expression of HOXB13 sequences may be used as an indicator of nonresponsive cases. The present invention provides a non-empirical means for determining responsiveness to TAM or another SERM in ER+ or ER- patients. This provides advantages over the use of a "wait and see" approach following treatment with TAM or other "antiestrogen" agent against breast cancer. The expression levels of these sequences may also be used as a means to assay small, node negative tumors that are not readily assessed by conventional means.

The expression levels of the identified sequences may be used alone or in combination with other sequences capable of determining responsiveness to treatment with TAM or another "antiestrogen" agent against breast cancer. Preferably, the sequences of the invention are used alone or in combination with each other, such as in the format of a ratio of expression levels that can have improved predictive power over analysis based on expression of sequences corresponding to individual genes. The invention provides for ratios of the expression level of a sequence that is

underexpressed to the expression level of a sequence that is overexpressed as an indicator of responsiveness or non-responsiveness.

The present invention provides means for correlating a molecular expression phenotype with a physiological response in a subject with ER+ or ER- breast cancer. This correlation provides a way to molecularly diagnose and/or determine treatment for a breast cancer afflicted subject. Additional uses of the sequences are in the classification of cells and tissues; and determination of diagnosis and/or prognosis. Use of the sequences to identify cells of a sample as responsive, or not, to treatment with TAM or other “antiestrogen” agent against breast cancer may be used to determine the choice, or alteration, of therapy used to treat such cells in the subject, as well as the subject itself, from which the sample originated.

Such methods of the invention may be used to assist the determination of providing tamoxifen or another “antiestrogen” agent against breast cancer as a chemopreventive or chemoprotective agent to a subject at high risk for development of breast cancer. These methods of the invention are an advance over the studies of Fabian et al. (J Natl Cancer Inst. 92(15):1217-27, 2000), which proposed a combination of cytomorphology and the Gail risk model to identify high risk patients. The methods may be used in combination with assessments of relative risk of breast cancer such as that discussed by Tan-Chiu et al. (J Natl Cancer Inst. 95(4):302-307, 2003). Non-limiting examples include assaying of minimally invasive sampling, such as random (periareolar) fine needle aspirates or ductal lavage samples (such as that described by Fabian et al. and optionally in combination with or as an addition to a mammogram positive for benign or malignant breast cancer), of breast cells for the expression levels of gene sequences as disclosed herein to assist in the determination of administering therapy with an “antiestrogen” agent against breast cancer, such as that which may occur in cases of high risk subjects (like those described by Tan-Chiu et al.). The assays would thus lead to the identification of subjects for who the application of an “antiestrogen” agent against breast cancer would likely be beneficial as a chemopreventive or chemoprotective agent. It is contemplated that such application as enabled by the instant invention could lead to beneficial effects such as those seen with the administration of tamoxifen (see for example, Wickerham D.L., *Breast Cancer Res. and Treatment* 75 Suppl 1:S7-12, Discussion S33-5, 2000). Other applications of the invention include assaying of advanced breast cancer, including metastatic

cancer, to determine the responsiveness, or non-responsiveness, thereof to treatment with an “antiestrogen” agent against breast cancer.

An assay of the invention may utilize a means related to the expression level of the sequences disclosed herein as long as the assay reflects, quantitatively or qualitatively, expression of the sequence. Preferably, however, a quantitative assay means is preferred. The ability to determine responsiveness to TAM or other “antiestrogen” agent against breast cancer and thus outcome of treatment therewith is provided by the recognition of the relevancy of the level of expression of the identified sequences and not by the form of the assay used to determine the actual level of expression. Identifying features of the sequences include, but are not limited to, unique nucleic acid sequences used to encode (DNA), or express (RNA), the disclosed sequences or epitopes specific to, or activities of, proteins encoded by the sequences. Alternative means include detection of nucleic acid amplification as indicative of increased expression levels and nucleic acid inactivation, deletion, or methylation, as indicative of decreased expression levels. Stated differently, the invention may be practiced by assaying one or more aspect of the DNA template(s) underlying the expression of the disclosed sequence(s), of the RNA used as an intermediate to express the sequence(s), or of the proteinaceous product expressed by the sequence(s), as well as proteolytic fragments of such products. As such, the detection of the presence of, amount of, stability of, or degradation (including rate) of, such DNA, RNA and proteinaceous molecules may be used in the practice of the invention.

The practice of the present invention is unaffected by the presence of minor mismatches between the disclosed sequences and those expressed by cells of a subject’s sample. A non-limiting example of the existence of such mismatches are seen in cases of sequence polymorphisms between individuals of a species, such as individual human patients within *Homo sapiens*. Knowledge that expression of the disclosed sequences (and sequences that vary due to minor mismatches) is correlated with the presence of non-normal or abnormal breast cells and breast cancer is sufficient for the practice of the invention with an appropriate cell containing sample via an assay for expression.

In one embodiment, the invention provides for the identification of the expression levels of the disclosed sequences by analysis of their expression in a sample containing ER+ or ER- breast

cells. In one preferred embodiment, the sample contains single cells or homogenous cell populations which have been dissected away from, or otherwise isolated or purified from, contaminating cells beyond that possible by a simple biopsy. Alternatively, undissected cells within a "section" of tissue may be used. Multiple means for such analysis are available, including
5 detection of expression within an assay for global, or near global, gene expression in a sample (e.g. as part of a gene expression profiling analysis such as on a microarray) or by specific detection, such as quantitative PCR (Q-PCR), or real time quantitative PCR.

Preferably, the sample is isolated via non-invasive or minimally invasive means. The expression of the disclosed sequence(s) in the sample may be determined and compared to the
10 expression of said sequence(s) in reference data of non-normal or cancerous breast cells. Alternatively, the expression level may be compared to expression levels in normal or non-cancerous cells, preferably from the same sample or subject. In embodiments of the invention utilizing Q-PCR, the expression level may be compared to expression levels of reference genes in the same sample or a ratio of expression levels may be used.

15 When individual breast cells are isolated in the practice of the invention, one benefit is that contaminating, non-breast cells (such as infiltrating lymphocytes or other immune system cells) are not present to possibly affect detection of expression of the disclosed sequence(s). Such contamination is present where a biopsy is used to generate gene expression profiles. However, analysis of differential gene expression and correlation to ER+ breast cancer outcomes with both
20 isolated and non-isolated samples, as described herein, increases the confidence level of the disclosed sequences as capable of having significant predictive power with either type of sample.

While the present invention is described mainly in the context of human breast cancer, it may be practiced in the context of breast cancer of any animal known to be potentially afflicted by breast cancer. Preferred animals for the application of the present invention are mammals,
25 particularly those important to agricultural applications (such as, but not limited to, cattle, sheep, horses, and other "farm animals"), animal models of breast cancer, and animals for human companionship (such as, but not limited to, dogs and cats).

The above aspects and embodiments of the invention may be applied equally with respect to use of more than one "antiestrogen" agent against breast cancer. In the case of a combination of

agents, any combination of more than one SERM, SERD, or AI may be used in place of TAM or another “antiestrogen” agent against breast cancer. Aromatase is an enzyme that provides a major source of estrogen in body tissues including the breast, liver, muscle and fat. Without being bound by theory, and solely provided to assist in a better understanding of the invention, AIs are

5 understood to function in a manner comparable to TAM and other “antiestrogen” agents against breast cancer, which are thought to act as antagonists of estrogen receptor in breast tissues and thus as against breast cancer. AIs may be either nonsteroidal or steroidal agents. Examples of the former, which inhibit aromatase via the heme prosthetic group) include, but are not limited to, anastrozole (arimidex), letrozole (femara), and vorozole (rivisor), which have been used or
10 contemplated as treatments for metastatic breast cancer. Examples of steroidal AIs, which inactivate aromatase, include, but are not limited to, exemestane (aromasin), androstenedione, and formestane (lentaron).

Other forms of therapy to reduce estrogen levels include surgical or chemical ovarian ablation. The former is physical removal of the ovaries while the latter is the use of agents to block
15 ovarian production of estrogen. One non-limiting example of the latter are agonists of gonadotropin releasing hormone (GnRH), such as goserelin (zoladex). Of course the instant invention may also be practiced with these therapies in place of treatment with one or more “antiestrogen” agent against breast cancer.

The invention disclosed herein is based in part on the performance of a genome-wide
20 microarray analysis of hormone receptor-positive invasive breast tumors from 60 patients treated with adjuvant tamoxifen alone, leading to the identification of a two-gene expression ratio that is highly predictive of clinical outcome. This expression ratio, which is readily adapted to PCR-based analysis of standard paraffin-embedded clinical specimens, was validated in an independent set of 20 patients as described below.

Brief Description of the Drawings

Figure 1 shows receiver operating characteristic (ROC) analyses of IL17BR, HOXB13, and CACNA1D expression levels as predictors of breast cancer outcomes in whole tissue sections (top 3 graphs) and laser microdissected cells (bottom 3 graphs). AUC refers to area under the curve.

5 Figure 2 contains six parts relating to the validation of a ratio of HOXB13 expression to IL17BR expression as an indicator of responsiveness, or lack thereof, to TAM. Parts a and b show the results of gene expression analysis of HOXB13 and IL17BR sequences by Q-PCR in both Responder and Non-responder samples. Plots of the Responder and Non-responder training and validation data sets are shown in Parts c and d, where “0” indicates Responder datapoints in both
10 and “1” indicates Non-responder datapoints in both. Parts e and f show plots of the Responder and Non-responder training and validation data sets as a function of survival, where the upper line in each Part represents the Responders and the lower line represents the Non-responders.

Modes of Practicing the Invention

Definitions of terms as used herein:

15 A gene expression “pattern” or “profile” or “signature” refers to the relative expression of genes correlated with responsiveness to treatment of ER+ breast cancer with TAM or another “antiestrogen” agent against breast cancer. Responsiveness or lack thereof may be expressed as survival outcomes which are correlated with an expression “pattern” or “profile” or “signature” that is able to distinguish between, and predict, said outcomes.

20 A “selective estrogen receptor modulator” or SERM is an “antiestrogen” agent that in some tissues act like estrogens (agonist) but block estrogen action in other tissues (antagonist). A “selective estrogen receptor downregulators” (or “SERD”s) or “pure” antiestrogens includes agents which block estrogen activity in all tissues. See Howell et al. (Best Practice & Res. Clin. Endocrinol. Metab. 18(1):47-66, 2004). Preferred SERMs of the invention are those that are
25 antagonists of estrogen in breast tissues and cells, including those of breast cancer. Non-limiting examples of such include TAM, raloxifene, GW5638, and ICI 182,780. The possible mechanisms of action by various SERMs have been reviewed (see for example Jordan et al., 2003, Breast Cancer Res. 5:281-283; Hall et al., 2001, J. Biol. Chem. 276(40):36869-36872; Dutertre et al. 2000, J.

Pharmacol. Exp. Therap. 295(2):431-437; and Wijayaratne et al., 1999, Endocrinology 140(12):5828-5840). Other non-limiting examples of SERMs in the context of the invention include triphenylethylenes, such as tamoxifen, GW5638, TAT-59, clomiphene, toremifene, droloxifene, and idoxifene; benzothiophenes, such as arzoxiphenes (LY353381 or LY353381-HCl);
5 benzopyrans, such as EM-800; naphthalenes, such as CP-336,156; and ERA-923.

Non-limiting examples of SERD or “pure” antiestrogens include agents such as ICI 182,780 (fulvestrant or faslodex) or the oral analogue SR16243 and ZK 191703 as well as aromatase inhibitors and chemical ovarian ablation agents as described herein.

Other agents encompassed by SERM as used herein include progesterone receptor inhibitors
10 and related drugs, such as progestomimetics like medroxyprogesterone acetate, megace, and RU-486; and peptide based inhibitors of ER action, such as LH-RH analogs (leuprolide, zoladex, [D-Trp6]LH-RH), somatostatin analogs, and LXXLL motif mimics of ER as well as tibolone and resveratrol. As noted above, preferred SERMs of the invention are those that are antagonist of estrogen in breast tissues and cells, including those of breast cancer. Non-limiting examples of
15 preferred SERMs include the actual or contemplated metabolites (in vivo) of any SERM, such as, but not limited to, 4-hydroxytamoxifen (metabolite of tamoxifen), EM652 (or SCH 57068 where EM-800 is a prodrug of EM-652), and GW7604 (metabolite of GW5638). See Willson et al. (1997, Endocrinology 138(9):3901-3911) and Dauvois et al. (1992, Proc. Nat’l. Acad. Sci., USA 89:4037-4041) for discussions of some specific SERMs.

20 Other preferred SERMs are those that produce the same relevant gene expression profile as tamoxifen or 4-hydroxytamoxifen. One example of means to identify such SERMs is provided by Levenson et al. (2002, Cancer Res. 62:4419-4426).

A “gene” is a polynucleotide that encodes a discrete product, whether RNA or proteinaceous in nature. It is appreciated that more than one polynucleotide may be capable of encoding a discrete
25 product. The term includes alleles and polymorphisms of a gene that encodes the same product, or a functionally associated (including gain, loss, or modulation of function) analog thereof, based upon chromosomal location and ability to recombine during normal mitosis.

A “sequence” or “gene sequence” as used herein is a nucleic acid molecule or polynucleotide composed of a discrete order of nucleotide bases. The term includes the ordering of

bases that encodes a discrete product (i.e. “coding region”), whether RNA or proteinaceous in nature, as well as the ordered bases that precede or follow a “coding region”. Non-limiting examples of the latter include 5’ and 3’ untranslated regions of a gene. It is appreciated that more than one polynucleotide may be capable of encoding a discrete product. It is also appreciated that alleles and polymorphisms of the disclosed sequences may exist and may be used in the practice of the invention to identify the expression level(s) of the disclosed sequences or the allele or polymorphism. Identification of an allele or polymorphism depends in part upon chromosomal location and ability to recombine during mitosis.

The terms “correlate” or “correlation” or equivalents thereof refer to an association between expression of one or more genes and a physiological response of a breast cancer cell and/or a breast cancer patient in comparison to the lack of the response. A gene may be expressed at higher or lower levels and still be correlated with responsiveness, non-responsiveness or breast cancer survival or outcome. The invention provides for the correlation between increases in expression of IL17BR and CACNA1D sequences and responsiveness of ER+ breast cells to TAM or another “antiestrogen” agent against breast cancer. Thus increases are indicative of responsiveness. Conversely, the lack of increases, including unchanged expression levels, are indicators of non-responsiveness. Similarly, the invention provides for the correlation between decreases in expression of HOXB13 sequences and responsiveness of ER+ breast cells to TAM or another SERM. Thus decreases are indicative of responsiveness while the lack of decreases, including unchanged expression levels, are indicators of non-responsiveness. Increases and decreases may be readily expressed in the form of a ratio between expression in a non-normal cell and a normal cell such that a ratio of one (1) indicates no difference while ratios of two (2) and one-half indicate twice as much, and half as much, expression in the non-normal cell versus the normal cell, respectively. Expression levels can be readily determined by quantitative methods as described below.

For example, increases in IL17BR, CACNA1D, or HOXB13 expression can be indicated by ratios of or about 1.1, of or about 1.2, of or about 1.3, of or about 1.4, of or about 1.5, of or about 1.6, of or about 1.7, of or about 1.8, of or about 1.9, of or about 2, of or about 2.5, of or about 3, of or about 3.5, of or about 4, of or about 4.5, of or about 5, of or about 5.5, of or about 6, of or about

6.5, of or about 7, of or about 7.5, of or about 8, of or about 8.5, of or about 9, of or about 9.5, of or about 10, of or about 15, of or about 20, of or about 30, of or about 40, of or about 50, of or about 60, of or about 70, of or about 80, of or about 90, of or about 100, of or about 150, of or about 200, of or about 300, of or about 400, of or about 500, of or about 600, of or about 700, of or about 800, of or about 900, or of or about 1000. A ratio of 2 is a 100% (or a two-fold) increase in expression. Decreases in IL17BR, CACNA1D, or HOXB13 expression can be indicated by ratios of or about 0.9, of or about 0.8, of or about 0.7, of or about 0.6, of or about 0.5, of or about 0.4, of or about 0.3, of or about 0.2, of or about 0.1, of or about 0.05, of or about 0.01, of or about 0.005, of or about 0.001, of or about 0.0005, of or about 0.0001, of or about 0.00005, of or about 0.00001, of or about 0.000005, or of or about 0.000001.

For a given phenotype, a ratio of the expression of a gene sequence expressed at increased levels in correlation with the phenotype to the expression of a gene sequence expressed at decreased levels in correlation with the phenotype may also be used as an indicator of the phenotype. As a non-limiting example, the phenotype of non-responsiveness to tamoxifen treatment of breast cancer is correlated with increased expression of HOXB13 as well as decreased expression of IL17BR and CACNA1D. Therefore, a ratio of the expression levels of HOXB13 to IL17BR (or CACNA1D) may be used as an indicator of non-responsiveness.

A “polynucleotide” is a polymeric form of nucleotides of any length, either ribonucleotides or deoxyribonucleotides. This term refers only to the primary structure of the molecule. Thus, this term includes double- and single-stranded DNA and RNA. It also includes known types of modifications including labels known in the art, methylation, “caps”, substitution of one or more of the naturally occurring nucleotides with an analog, and internucleotide modifications such as uncharged linkages (e.g., phosphorothioates, phosphorodithioates, etc.), as well as unmodified forms of the polynucleotide.

The term “amplify” is used in the broad sense to mean creating an amplification product can be made enzymatically with DNA or RNA polymerases. “Amplification,” as used herein, generally refers to the process of producing multiple copies of a desired sequence, particularly those of a sample. “Multiple copies” mean at least 2 copies. A “copy” does not necessarily mean perfect sequence complementarity or identity to the template sequence. Methods for amplifying mRNA are

generally known in the art, and include reverse transcription PCR (RT-PCR) and those described in U.S. Patent Application 10/062,857 (filed on October 25, 2001), as well as U.S. Provisional Patent Applications 60/298,847 (filed June 15, 2001) and 60/257,801 (filed December 22, 2000), all of which are hereby incorporated by reference in their entireties as if fully set forth. Another method which may be used is quantitative PCR (or Q-PCR). Alternatively, RNA may be directly labeled as the corresponding cDNA by methods known in the art.

By "corresponding", it is meant that a nucleic acid molecule shares a substantial amount of sequence identity with another nucleic acid molecule. Substantial amount means at least 95%, usually at least 98% and more usually at least 99%, and sequence identity is determined using the BLAST algorithm, as described in Altschul et al. (1990), J. Mol. Biol. 215:403-410 (using the published default setting, i.e. parameters $w=4$, $t=17$).

A "microarray" is a linear or two-dimensional or three dimensional (and solid phase) array of preferably discrete regions, each having a defined area, formed on the surface of a solid support such as, but not limited to, glass, plastic, or synthetic membrane. The density of the discrete regions on a microarray is determined by the total numbers of immobilized polynucleotides to be detected on the surface of a single solid phase support, preferably at least about $50/\text{cm}^2$, more preferably at least about $100/\text{cm}^2$, even more preferably at least about $500/\text{cm}^2$, but preferably below about $1,000/\text{cm}^2$. Preferably, the arrays contain less than about 500, about 1000, about 1500, about 2000, about 2500, or about 3000 immobilized polynucleotides in total. As used herein, a DNA microarray is an array of oligonucleotides or polynucleotides placed on a chip or other surfaces used to hybridize to amplified or cloned polynucleotides from a sample. Since the position of each particular group of primers in the array is known, the identities of a sample polynucleotides can be determined based on their binding to a particular position in the microarray. As an alternative to the use of a microarray, an array of any size may be used in the practice of the invention, including an arrangement of one or more position of a two-dimensional or three dimensional arrangement in a solid phase to detect expression of a single gene sequence.

Because the invention relies upon the identification of genes that are over- or under-expressed, one embodiment of the invention involves determining expression by hybridization of mRNA, or an amplified or cloned version thereof, of a sample cell to a polynucleotide that is unique

to a particular gene sequence. Preferred polynucleotides of this type contain at least about 16, at least about 18, at least about 20, at least about 22, at least about 24, at least about 26, at least about 28, at least about 30, or at least about 32 consecutive basepairs of a gene sequence that is not found in other gene sequences. The term "about" as used in the previous sentence refers to an increase or decrease of 1 from the stated numerical value. Even more preferred are polynucleotides of at least or about 50, at least or about 100, at least about or 150, at least or about 200, at least or about 250, at least or about 300, at least or about 350, at least or about 400, , at least or about 450, or at least or about 500 consecutive bases of a sequence that is not found in other gene sequences. The term "about" as used in the preceding sentence refers to an increase or decrease of 10% from the stated numerical value. Longer polynucleotides may of course contain minor mismatches (e.g. via the presence of mutations) which do not affect hybridization to the nucleic acids of a sample. Such polynucleotides may also be referred to as polynucleotide probes that are capable of hybridizing to sequences of the genes, or unique portions thereof, described herein. Such polynucleotides may be labeled to assist in their detection. Preferably, the sequences are those of mRNA encoded by the genes, the corresponding cDNA to such mRNAs, and/or amplified versions of such sequences. In preferred embodiments of the invention, the polynucleotide probes are immobilized on an array, other solid support devices, or in individual spots that localize the probes.

In another embodiment of the invention, all or part of a disclosed sequence may be amplified and detected by methods such as the polymerase chain reaction (PCR) and variations thereof, such as, but not limited to, quantitative PCR (Q-PCR), reverse transcription PCR (RT-PCR), and real-time PCR (including as a means of measuring the initial amounts of mRNA copies for each sequence in a sample), optionally real-time RT-PCR or real-time Q-PCR. Such methods would utilize one or two primers that are complementary to portions of a disclosed sequence, where the primers are used to prime nucleic acid synthesis. The newly synthesized nucleic acids are optionally labeled and may be detected directly or by hybridization to a polynucleotide of the invention. The newly synthesized nucleic acids may be contacted with polynucleotides (containing sequences) of the invention under conditions which allow for their hybridization. Additional methods to detect the expression of expressed nucleic acids include RNase protection assays, including liquid phase hybridizations, and *in situ* hybridization of cells.

Alternatively, and in yet another embodiment of the invention, gene expression may be determined by analysis of expressed protein in a cell sample of interest by use of one or more antibodies specific for one or more epitopes of individual gene products (proteins), or proteolytic fragments thereof, in said cell sample or in a bodily fluid of a subject. The cell sample may be one of breast cancer epithelial cells enriched from the blood of a subject, such as by use of labeled antibodies against cell surface markers followed by fluorescence activated cell sorting (FACS). Such antibodies are preferably labeled to permit their easy detection after binding to the gene product. Detection methodologies suitable for use in the practice of the invention include, but are not limited to, immunohistochemistry of cell containing samples or tissue, enzyme linked immunosorbent assays (ELISAs) including antibody sandwich assays of cell containing tissues or blood samples, mass spectroscopy, and immuno-PCR.

The term “label” refers to a composition capable of producing a detectable signal indicative of the presence of the labeled molecule. Suitable labels include radioisotopes, nucleotide chromophores, enzymes, substrates, fluorescent molecules, chemiluminescent moieties, magnetic particles, bioluminescent moieties, and the like. As such, a label is any composition detectable by spectroscopic, photochemical, biochemical, immunochemical, electrical, optical or chemical means.

The term “support” refers to conventional supports such as beads, particles, dipsticks, fibers, filters, membranes and silane or silicate supports such as glass slides.

As used herein, a “breast tissue sample” or “breast cell sample” refers to a sample of breast tissue or fluid isolated from an individual suspected of being afflicted with, or at risk of developing, breast cancer. Such samples are primary isolates (in contrast to cultured cells) and may be collected by any non-invasive or minimally invasive means, including, but not limited to, ductal lavage, fine needle aspiration, needle biopsy, the devices and methods described in U.S. Patent 6,328,709, or any other suitable means recognized in the art. Alternatively, the “sample” may be collected by an invasive method, including, but not limited to, surgical biopsy.

“Expression” and “gene expression” include transcription and/or translation of nucleic acid material.

As used herein, the term “comprising” and its cognates are used in their inclusive sense; that is, equivalent to the term “including” and its corresponding cognates.

Conditions that “allow” an event to occur or conditions that are “suitable” for an event to occur, such as hybridization, strand extension, and the like, or “suitable” conditions are conditions that do not prevent such events from occurring. Thus, these conditions permit, enhance, facilitate, and/or are conducive to the event. Such conditions, known in the art and described herein, depend upon, for example, the nature of the nucleotide sequence, temperature, and buffer conditions. These conditions also depend on what event is desired, such as hybridization, cleavage, strand extension or transcription.

Sequence “mutation,” as used herein, refers to any sequence alteration in the sequence of a gene disclosed herein interest in comparison to a reference sequence. A sequence mutation includes single nucleotide changes, or alterations of more than one nucleotide in a sequence, due to mechanisms such as substitution, deletion or insertion. Single nucleotide polymorphism (SNP) is also a sequence mutation as used herein. Because the present invention is based on the relative level of gene expression, mutations in non-coding regions of genes as disclosed herein may also be assayed in the practice of the invention.

“Detection” includes any means of detecting, including direct and indirect detection of gene expression and changes therein. For example, “detectably less” products may be observed directly or indirectly, and the term indicates any reduction (including the absence of detectable signal). Similarly, “detectably more” product means any increase, whether observed directly or indirectly.

Increases and decreases in expression of the disclosed sequences are defined in the following terms based upon percent or fold changes over expression in normal cells. Increases may be of 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 120, 140, 160, 180, or 200% relative to expression levels in normal cells. Alternatively, fold increases may be of 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5, 9, 9.5, or 10 fold over expression levels in normal cells. Decreases may be of 10, 20, 30, 40, 50, 55, 60, 65, 70, 75, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 99 or 100% relative to expression levels in normal cells.

Unless defined otherwise all technical and scientific terms used herein have the same meaning as commonly understood to one of ordinary skill in the art to which this invention belongs.

Embodiments of the Invention

In a first aspect, the disclosed invention relates to the identification and use of gene expression patterns (or profiles or “signatures”) which discriminate between (or are correlated with) breast cancer survival in a subject treated with tamoxifen (TAM) or another “antiestrogen” agent against breast cancer. Such patterns may be determined by the methods of the invention by use of a number of reference cell or tissue samples, such as those reviewed by a pathologist of ordinary skill in the pathology of breast cancer, which reflect breast cancer cells as opposed to normal or other non-cancerous cells. The outcomes experienced by the subjects from whom the samples may be correlated with expression data to identify patterns that correlate with the outcomes following treatment with TAM or another “antiestrogen” agent against breast cancer. Because the overall gene expression profile differs from person to person, cancer to cancer, and cancer cell to cancer cell, correlations between certain cells and genes expressed or underexpressed may be made as disclosed herein to identify genes that are capable of discriminating between breast cancer outcomes.

The present invention may be practiced with any number of the genes believed, or likely to be, differentially expressed with respect to breast cancer outcomes, particularly in cases of ER+ breast cancer. The identification may be made by using expression profiles of various homogenous breast cancer cell populations, which were isolated by microdissection, such as, but not limited to, laser capture microdissection (LCM) of 100-1000 cells. The expression level of each gene of the expression profile may be correlated with a particular outcome. Alternatively, the expression levels of multiple genes may be clustered to identify correlations with particular outcomes.

Genes with significant correlations to breast cancer survival when the subject is treated with tamoxifen may be used to generate models of gene expressions that would maximally discriminate between outcomes where a subject responds to treatment with tamoxifen or another “antiestrogen” agent against breast cancer and outcomes where the treatment is not successful. Alternatively, genes with significant correlations may be used in combination with genes with lower correlations without significant loss of ability to discriminate between outcomes. Such models may be generated by any appropriate means recognized in the art, including, but not limited to, cluster analysis, supported vector machines, neural networks or other algorithm known in the art. The

models are capable of predicting the classification of a unknown sample based upon the expression of the genes used for discrimination in the models. "Leave one out" cross-validation may be used to test the performance of various models and to help identify weights (genes) that are uninformative or detrimental to the predictive ability of the models. Cross-validation may also be used to identify genes that enhance the predictive ability of the models.

The gene(s) identified as correlated with particular breast cancer outcomes relating to tamoxifen treatment by the above models provide the ability to focus gene expression analysis to only those genes that contribute to the ability to identify a subject as likely to have a particular outcome relative to another. The expression of other genes in a breast cancer cell would be relatively unable to provide information concerning, and thus assist in the discrimination of, a breast cancer outcome.

As will be appreciated by those skilled in the art, the models are highly useful with even a small set of reference gene expression data and can become increasingly accurate with the inclusion of more reference data although the incremental increase in accuracy will likely diminish with each additional datum. The preparation of additional reference gene expression data using genes identified and disclosed herein for discriminating between different outcomes in breast cancer following treatment with tamoxifen or another "antiestrogen" agent against breast cancer is routine and may be readily performed by the skilled artisan to permit the generation of models as described above to predict the status of an unknown sample based upon the expression levels of those genes.

To determine the (increased or decreased) expression levels of genes in the practice of the present invention, any method known in the art may be utilized. In one preferred embodiment of the invention, expression based on detection of RNA which hybridizes to the genes identified and disclosed herein is used. This is readily performed by any RNA detection or amplification+detection method known or recognized as equivalent in the art such as, but not limited to, reverse transcription-PCR, the methods disclosed in U.S. Patent Application 10/062,857 (filed on October 25, 2001) as well as U.S. Provisional Patent Applications 60/298,847 (filed June 15, 2001) and 60/257,801 (filed December 22, 2000), and methods to detect the presence, or absence, of RNA stabilizing or destabilizing sequences.

Alternatively, expression based on detection of DNA status may be used. Detection of the DNA of an identified gene as methylated or deleted may be used for genes that have decreased expression in correlation with a particular breast cancer outcome. This may be readily performed by PCR based methods known in the art, including, but not limited to, Q-PCR. Conversely, detection of the DNA of an identified gene as amplified may be used for genes that have increased expression in correlation with a particular breast cancer outcome. This may be readily performed by PCR based, fluorescent *in situ* hybridization (FISH) and chromosome *in situ* hybridization (CISH) methods known in the art.

Expression based on detection of a presence, increase, or decrease in protein levels or activity may also be used. Detection may be performed by any immunohistochemistry (IHC) based, blood based (especially for secreted proteins), antibody (including autoantibodies against the protein) based, exfoliate cell (from the cancer) based, mass spectroscopy based, and image (including used of labeled ligand) based method known in the art and recognized as appropriate for the detection of the protein. Antibody and image based methods are additionally useful for the localization of tumors after determination of cancer by use of cells obtained by a non-invasive procedure (such as ductal lavage or fine needle aspiration), where the source of the cancerous cells is not known. A labeled antibody or ligand may be used to localize the carcinoma(s) within a patient or to assist in the enrichment of exfoliated cancer cells from a bodily fluid.

A preferred embodiment using a nucleic acid based assay to determine expression is by immobilization of one or more sequences of the genes identified herein on a solid support, including, but not limited to, a solid substrate as an array or to beads or bead based technology as known in the art. Alternatively, solution based expression assays known in the art may also be used. The immobilized gene(s) may be in the form of polynucleotides that are unique or otherwise specific to the gene(s) such that the polynucleotide would be capable of hybridizing to a DNA or RNA corresponding to the gene(s). These polynucleotides may be the full length of the gene(s) or be short sequences of the genes (up to one nucleotide shorter than the full length sequence known in the art by deletion from the 5' or 3' end of the sequence) that are optionally minimally interrupted (such as by mismatches or inserted non-complementary basepairs) such that hybridization with a DNA or RNA corresponding to the gene(s) is not affected. Preferably, the polynucleotides used are

from the 3' end of the gene, such as within about 350, about 300, about 250, about 200, about 150, about 100, or about 50 nucleotides from the polyadenylation signal or polyadenylation site of a gene or expressed sequence. Polynucleotides containing mutations relative to the sequences of the disclosed genes may also be used so long as the presence of the mutations still allows hybridization to produce a detectable signal.

The immobilized gene(s) may be used to determine the state of nucleic acid samples prepared from sample breast cell(s) for which the outcome of the sample's subject (e.g. patient from whom the sample is obtained) is not known or for confirmation of an outcome that is already assigned to the sample's subject. Without limiting the invention, such a cell may be from a patient with ER+ or ER- breast cancer. The immobilized polynucleotide(s) need only be sufficient to specifically hybridize to the corresponding nucleic acid molecules derived from the sample under suitable conditions. While even a single correlated gene sequence may be able to provide adequate accuracy in discriminating between two breast cancer outcomes, two or more, three or more, four or more, five or more, six or more, seven or more, eight or more, nine or more, ten or more, or eleven or more of the genes identified herein may be used as a subset capable of discriminating may be used in combination to increase the accuracy of the method. The invention specifically contemplates the selection of more than one, two or more, three or more, four or more, five or more, six or more, seven or more, eight or more, nine or more, ten or more, or eleven or more of the genes disclosed in the tables and figures herein for use as a subset in the identification of breast cancer survival outcome.

Of course 1 or more, 2 or more, 3 or more, 4 or more, 5 or more, 6 or more, 7 or more, 8 or more, 9 or more, or all the genes provided in Tables 2 and/or 3 below may be used. "Accession" as used in the context of the Tables herein as well as the present invention refers to the GenBank accession number of a sequence of each gene, the sequences of which are hereby incorporated by reference in their entireties as they are available from GenBank as accessed on the filing date of the present application. P value refers to values assigned as described in the Examples below. The indications of "E-xx" where "xx" is a two digit number refers to alternative notation for exponential figures where "E-xx" is " 10^{-xx} ". Thus in combination with the numbers to the left of "E-xx", the

value being represented is the numbers to the left times 10^{-xx} . "Description" as used in the Tables provides a brief identifier of what the sequence/gene encodes.

Genes with a correlation identified by a p value below or about 0.02, below or about 0.01, below or about 0.005, or below or about 0.001 are preferred for use in the practice of the invention.

5 The present invention includes the use of gene(s) the expression of which identify different breast cancer outcomes after treatment with TAM or another "antiestrogen" agent against breast cancer to permit simultaneous identification of breast cancer survival outcome of a patient based upon assaying a breast cancer sample from said patient.

10 In a second aspect, the present invention relates to the identification and use of three sets of sequences for the determination of responsiveness of ER+ breast cancer to treatment with TAM or another "antiestrogen" agent against breast cancer. The differential expression of these sequences in breast cancer relative to normal breast cells is used to predict responsiveness to TAM or another "antiestrogen" agent against breast cancer in a subject.

15 To identify gene expression patterns in ER positive, early stage invasive breast cancers that might predict response to hormonal therapy, microarray gene expression analysis was performed on tumors from 60 women uniformly treated with adjuvant tamoxifen alone. These patients were identified from a total of 103 ER+ early stage cases presenting to Massachusetts General Hospital between 1987 and 1997, from whom tumor specimens were snap frozen and for whom minimal 5 year follow-up was available (see Table 1 for details). Within this cohort, 28 (46%) women
20 developed distant metastasis with a median time to recurrence of 4 years ("tamoxifen non-responders") and 32 (54%) women remained disease-free with median follow-up of 10 years ("tamoxifen responders"). Responders were matched with non-responder cases with respect to TNM staging (see Singletary, S.E. et al. "Revision of the American Joint Committee on Cancer staging system for breast cancer." J Clin Oncol 20, 3628-36 (2002)) and tumor grade (see Dalton,
25 L.W. et al. "Histologic grading of breast cancer: linkage of patient outcome with level of pathologist agreement." Mod Pathol 13, 730-5. (2000)).

Previous studies linking gene expression profiles to clinical outcome in breast cancer have demonstrated that the potential for distant metastasis and overall survival probability may be predictable through biological characteristics of the primary tumor at the time of diagnosis (see

Huang, E. et al. "Gene expression predictors of breast cancer outcomes." Lancet 361, 1590-6 (2003); Sorlie, T. et al. "Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications." Proc Natl Acad Sci U S A 98:10869-74 (2001); Sorlie, T. et al. "Repeated observation of breast tumor subtypes in independent gene expression data sets." Proc Natl Acad Sci U S A 100, 8418-23 (2003); Sotiriou, C. et al. "Breast cancer classification and prognosis based on gene expression profiles from a population-based study." Proc Natl Acad Sci U S A 100, 10393-8 (2003); van 't Veer, L.J. et al. "Gene expression profiling predicts clinical outcome of breast cancer." Nature 415, 530-6 (2002); and van de Vijver, M.J. et al. "A gene-expression signature as a predictor of survival in breast cancer." N Engl J Med 347, 1999-2009 (2002)). In particular, a 70-gene expression signature has proven to be a strong prognostic factor, out-performing all known clinicopathological parameters. However, in those studies patients either received no adjuvant therapy (van 't Veer, L.J. et al. Nature 2002) or were treated non-uniformly with hormonal and chemotherapeutic regimens (Huang, E. et al.; Sorlie, T. et al.; Sorlie, T. et al.; Sotiriou, C. et al.; and van de Vijver, M.J. et al. N Engl J Med 2002). Patients with ER+ early-stage breast cancer treated with tamoxifen alone, such as the cohort studied here, represent only a subset of the population tested with the 70-gene signature. Of note, 61 of the genes in the 70-gene signature were present on the microarray used as described below, but no significant association with clinical outcome was observed in the defined subset of patients.

In comparison with existing biomarkers, including ESR1, PGR, ERBB2 and EGFR, three sets of gene sequences disclosed herein are significantly more predictive of responsiveness to TAM treatment. Multivariate analysis indicated that these three genes were significant predictors of clinical outcome independent of tumor size, nodal status and tumor grade. ER and progesterone receptor (PR) expression have been the major clinicopathological predictors for response to TAM. However, up to 40% of ER+ tumors fail to respond or develop resistance to TAM. The invention thus provides for the use of the identified biomarkers to allow better patient management by identifying patients who are more likely to benefit from TAM or other endocrine therapy and those who are likely to develop resistance and tumor recurrence.

As noted herein, the sequences(s) identified by the present invention are expressed in correlation with ER+ breast cancer cells. For example, IL17BR, identified by I.M.A.G.E.

Consortium Clusters NM_018725 and NM_172234 ("The I.M.A.G.E. Consortium: An Integrated Molecular Analysis of Genomes and their Expression," Lennon et al., 1996, Genomics 33:151-152; see also image.llnl.gov) has been found to be useful in predicting responsiveness to TAM treatment.

In preferred embodiments of the invention, any sequence, or unique portion thereof, of the
5 IL17BR sequences of the cluster, as well as the UniGene *Homo sapiens* cluster Hs.5470, may be used. Similarly, any sequence encoding all or a part of the protein encoded by any IL17BR sequence disclosed herein may be used. Consensus sequences of I.M.A.G.E. Consortium clusters are as follows, with the assigned coding region (ending with a termination codon) underlined and preceded by the 5' untranslated and/or non-coding region and followed by the 3' untranslated
10 and/or non-coding region:

SEQ ID NO:1 (consensus sequence for IL17BR, transcript variant 1, identified as NM_018725 or NM_018725.2)

15 agcgcagcgt gcggttgccc tggatccgc gcagtggccc ggcatgtcg ctctgtctgc
taagcctggc cgctgtgtgc aggagcgccg taccgagaga gccgaccgtt caatgtggct
ctgaaactgg gccatctcca gagtggatgc tacaacatga tctaattccc ggagacttga
gggacctccg agtagaacct gttacaacta gtgttgcaac aggggactat tcaattttga
tgaatgtaag ctgggtactc cgggcagatg ccagcatccg cttgttgaag gccaccaaga
20 tttgtgtgac gggcaaaagc aacttcagc cctacagctg tgtgaggtgc aattacacag
aggccttcca gactcagacc agaccctctg gtggtaaatg gacattttcc tacatcggct
tccctgtaga gctgaacaca gtctatttca ttggggccca taatattcct aatgcaaata
tgaatgaaga tggcccttcc atgtctgtga atttcacctc accaggctgc ctagaccaca
taatgaaata taaaaaaaag tgtgtcaagg ccggaagcct gtgggatccg aacatcactg
25 cttgtaagaa gaatgaggag acagtagaag tgaacttcac aaccactccc ctgggaagaa
gatacatggc tcttatccaa cacagcacta tcatcgggtt ttctcaggtg tttgagccac
accagaagaa acaaacgcga gcttcagtggtg tgaattccagt gactggggat agtgaagggtg
ctacgggtgca gctgactcca tattttcccta cttgtggcag cgactgcata cgacataaag
gaacagttgt gctctgcccc caaacaggcg tccctttccc tctggataac aacaaaagca
30 agccgggagg ctggctgcct ctctctctgc tgtctctgct ggtggccaca tgggtgctgg
tggcagggat ctatctaagtg tggaggcacg aaaggatcaa gaagacttcc ttttctacca
ccacactact gccccccatt aagggttcttg tggtttacc ctttgaaata tgtttccatc
acacaatttg ttacttcact gaatttcttc aaaaccattg cagaagtgaag gtcactcctg
aaaagtggca gaaaaagaaa atagcagaga tgggtccagt gcagtggctt gccactcaaa
35 agaaggcagc agacaaagtc gtcttccttc tttccaatga cgtcaacagt gtgtgcatg
gtacctgtgg caagagcgag ggcagtccca gtgagaactc tcaagacctc tcccccttg
cctttaacct tttctgcagt gatctaagaa gccagattca tctgcacaaa tacgtgggtg
tctactttag agagattgat acaaaagacg attacaatgc tctcagtgtc tgccccaagt

```

accacctcat gaaggatgcc actgctttct gtgcagaact tctccatgtc aagcagcagg
tgtcagcagg aaaaagatca caagcctgcc acgatggctg ctgctccttg tagcccaccc
atgagaagca agagacctta aaggcttcct atcccaccaa ttacaggga aaaacgtgtg
atgatcctga agcttactat gcagcctaca aacagcctta gtaattaaaa cattttatac
5 caataaaatt ttcaaatatt gctaactaat gtagcattaa ctaacgattg gaaactacat
ttacaacttc aaagctgttt tatacataga aatcaattac agttttaatt gaaaactata
accattttga taatgcaaca ataaagcatc ttcagccaaa catctagtct tccatagacc
atgcattgca gtgtacccag aactgttttag ctaatatctt atgtttaatt aatgaatact
aactctaaga acccctcact gattcactca atagcatctt aagtgaaaaa ccttctatta
10 catgcaaaaa atcattgttt ttaagataac aaaagtaggg aataaacaag ctgaaccac
ttttaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaa

```

SEQ ID NO:2 (consensus sequence for IL17BR, transcript variant 2, identified as NM_172234 or NM_172234.1)

```

15
agcgcagcgt gcgggtggcc tggatcccg gcagtggccc ggcgatgtcg ctcggtgtgc
taagcctggc cgcgctgtgc aggagcgccg taccgagaga gccgaccgtt caatgtggct
ctgaaactgg gccatctcca gagtggatgc tacaacatga tctaataccc ggagacttga
gggacctccg agtagaacct gttacaacta gtgttgcaac aggggactat tcaattttga
20 tgaatgtaag ctgggtactc cgggcagatg ccagcatccg cttgttgaaag gccaccaaga
tttgtgtgac gggcaaaagc aacttccagt cctacagctg tgtgaggtgc aattacacag
aggccttcca gactcagacc agaccctctg gtggtaaagt gacattttcc tacatcggct
tccctgtaga gctgaacaca gtctatttca ttggggccca taatattcct aatgcaaata
tgaatgaaga tggcccttcc atgtctgtga atttcacctc accaggctgc ctagaccaca
25 taatgaaata taaaaaaaaag tgtgtcaagg ccggaagcct gtgggatccg aacatcactg
cttgtaagaa gaatgaggag acagtagaag tgaacttcac aaccactccc ctgggaacaa
gatacatggc tcttatccaa cacagcacta tcatcggtt ttctcaggtg tttgagccac
accagaagaa acaaacgcga gcttcagtgg tgattccagt gactggggat agtgaagggtg
ctacggtgca ggtaaagtcc agtgagctgc tctggggagg gaaggacat agaagactgt
30 tccatcattc attgctttta aggatgagtt ctctcttgtc aaatgcactt ctgccagcag
acaccagtta agtggcggtc atgggggctc tttcgctgca gcctccaccg tgctgaggtc
aggaggccga cgtggcagtt gtgggtccctt ttgcttgat taatggctgc tgaccttcca
aagcactttt tattttcatt ttctgtcaca gacactcagg gatagcagta ccattttact
tccgcaagcc tttaactgca agatgaagct gcaaagggtt tgaaatggga aggtttgagt
35 tccaggcagc gtatgaactc tggagagggg ctgccagtc tctctggggc gcagcggacc
cagctggaac acaggaagtt ggagcagtag gtgctccttc acctctcagt atgtctcttt
caactctagt ttttgaggtg gggacacagg aggtccagt ggacacagcc actcccaaa
gagtaaggag cttccatgct tcattccctg gcataaaaag tgctcaaaca caccagaggg
ggcaggcacc agccagggtg tgatggctac tacccttttc tggagaacca tagacttccc
40 ttactacagg gacttgcatt tcctaaagca ctggctgaag gaagccaaga ggatcactgc
tgctcctttt ttctagagga aatgtttgtc tacgtggtaa gatatgacct agccctttta
ggtaagcgaa ctggtatgtt agtaacgtgt acaaagttaa ggttcagacc ccgggagttc
tgggcacgtg ggtctcgggt cactggtttt gacttttaggg ctttggtaca gatgtgtgac
caaggggaaa atgtgcatga caacactaga ggtatgggcg aagccagaaa gaagggaagt

```

```

    tttggctgaa gtaggagtct tgggtgagatt ttgctctgat gcatgggtgtg aactttctga
    gcctcttggt tttcctcagc tgactccata ttttctact tgtggcagcg actgcatccg
    acataaagga acagttgtgc tctgcccaca aacaggcgctc cttttccctc tggataacaa
    caaaagcaag cggggaggct ggctgcctct cctcctgctg tctctgctgg tggccacatg
5   ggtgctggtg gcagggatct atctaattgtg gaggcacgaa aggatcaaga agacttcctt
    ttctaccacc acactactgc cccccattaa ggttcttggtg gtttaccat ctgaaatatg
    tttccatcac acaatttggt acttcactga atttcttcaa aaccattgca gaagtggagt
    catccttgaa aagtggcaga aaaagaaaat agcagagatg ggtccagtgc agtggcttgc
    cactcaaaag aaggcagcag acaaagtcgt ctctcttctt tccaatgacg tcaacagtgt
10  gtgcatggt acctgtggca agagcgaggg cagtcccagt gagaactctc aagacctctt
    cccccttgcc tttaaccttt tctgcagtga tctaagaagc cagattcatc tgcacaaata
    cgtgggtggtc tacttttagag agattgatac aaaagacgat tacaatgctc tcagtgtctg
    ccccaagtac cacctcatga aggatgccac tgctttctgt gcagaacttc tccatgtcaa
    gcagcagggtg tcagcaggaa aaagatcaca agcctgccac gatggctgct gctccttgta
15  gccacccat gagaagcaag agacctaaa ggcttcctat cccaccaatt acagggaaaa
    aacgtgtgat gatcctgaag ctactatgc agctacaaa cagccttagt aattaaaaca
    ttttatacca ataaaatttt caaatattgc taactaatgt agcattaact aacgattgga
    aactacattt acaacttcaa agctgtttta tacatagaaa tcaattacag ttttaattga
    aaactataac cattttgata atgcaacaat aaagcatctt cagccaaaca tctagtcttc
20  catagaccat gcattgcagt gtaccagaa ctgttttagct aatattctat gtttaattaa
    tgaatactaa ctctaagaac ccctcactga ttcactcaat agcatcttaa gtgaaaaacc
    ttctattaca tgcaaaaaat cattgttttt aagataacaa aagtagggaa taaacaagct
    gaaccactt ttaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaa

```

25 I.M.A.G.E. Consortium Clone ID numbers and the corresponding GenBank accession
 numbers of sequences identified as belonging to the I.M.A.G.E. Consortium and UniGene clusters,
 are listed below. Also included are sequences that are not identified as having a Clone ID number
 but still identified as being those of IL17BR. The sequences include those of the "sense" and
 complementary strands sequences corresponding to IL17BR. The sequence of each GenBank
 30 accession number is presented in the attached Appendix.

Clone ID numbers	GenBank accession numbers
2985728	AW675096, AW673932, BC000980
5286745	BI602183
5278067	BI458542
5182255	BI823321
924000	AA514396
3566736	BF110326

3195409	BE466508
3576775	BF740045
2772915	AW299271
1368826	AA836217
1744837	AI203628
2285564	AI627783
2217709	AI744263
2103651	AI401622
2419487	AI826949
3125592	BE047352
2284721	AI911549
3643302	BF194822
1646910	AI034244
1647001	AI033911
3323709	BF064177
1419779	AA847767
2205190	AI538624
2295838	AI913613
2461335	AI942234
2130362	AI580483
2385555	AI831909
2283817	AI672344
2525596	AW025192
454687	AA677205
1285273	AA721647
3134106	BF115018
342259	W61238, W61239
1651991	AI032064
2687714	AW236941
3302808	BG057174
2544461	AW058532
122014	T98360, T98361
2139250	AI470845

2133899	AI497731
121300	T96629, T96740
162274	H25975, H25941
3446667	BE539514, BX282554
156864	R74038, R74129
4611491	BG433769
4697316	BG530489
429376	AA007528, AA007529
5112415	BI260259
701357	AA287951, AA287911
121909	T97852, T97745
268037	N40294
1307489	AA809841
1357543	AA832389
48442	H14692
1302619	AA732635
1562857	AA928257
1731938	AI184427
1896025	AI298577
2336350	AI692717
1520997	AA910922
240506	H90761
2258560	AI620122
1569921	AI793318, AA962325, AI733290
6064627	BQ226353
299018	W04890
5500181	BM455231
2484011	BI492426
4746376	BG674622
233783	BX111256
1569921	BX117618
450450	AA682806
1943085	AI202376

2250390	AI658949
4526156	BG403405
3249181	BE673417
2484395	AW021469
30515867	CF455736
2878155	AW339874
4556884	BG399724
3254505	BF475787
3650593	BF437145
233783	H64601
None (mRNA sequences)	AF212365, AF208110, AF208111, AF250309, AK095091
None	BM983744, CB305764, BM715988, BM670929, BI792416, BI715216, N56060, CB241389, AV660618, BX088671, CB154426, CA434589, CA412162, CA314073, BF921554, BF920093, AV685699, AV650175, BX483104, CD675121, BE081436, AW970151, AW837146, AW368264, D25960, AV709899, BX431018, AL535617, AL525465, BX453536, BX453537, AV728945, AV728939, AV727345

In one preferred embodiment, any sequence, or unique portion thereof, of the following IL17BR sequence, identified by AF208111 or AF208111.1, may be used in the practice of the invention.

5

SEQ ID NO:3 (sequence for IL17BR):

CGGCGATGTCGCTCGTGCTGATAAGCCTGGCCGCGCTGTGCAGGAGCGCCGTACCCCGAG
 AGCCGACCGTTCAATGTGGCTCTGAACTGGGCCATCTCCAGAGTGGATGCTACAACATG
 10 ATCTAATCCCGGAGACTTGAGGGACCTCCGAGTAGAACCTGTTACAACAGTGTGCAA
 CAGGGGACTATTCAATTTTGATGAATGTAAGCTGGGTACTCCGGGCAGATGCCAGCATCC
 GCTTGTTGAAGGCCACCAAGATTTGTGTGACGGGCAAAAGCAACTTCCAGTCCTACAGCT
 GTGTGAGGTGCAATTACACAGAGGCCTTCCAGACTCAGACCAGACCCTCTGGTGGTAAAT
 GGACATTTTCCTATATCGGCTTCCCTGTAGAGCTGAACACAGTCTATTTTCATTGGGGCCC
 15 ATAATATTCCTAATGCAAATATGAATGAAGATGGCCCTTCCATGTCTGTGAATTTACCT
 CACCAGGCTGCCTAGACCACATAATGAAATATAAAAAAAGTGTGTCAAGGCCGGAAGCC

TGTGGGATCCGAACATCACTGCTTGTAAGAAGAATGAGGAGACAGTAGAAGTGAACCTCA
CAACCACTCCCCTGGGAAACAGATACATGGCTCTTATCCAACACAGCACTATCATCGGGT
TTTCTCAGGTGTTTGAGCCACACCAGAAGAAACAAACGCGAGCTTCAGTGGTGATTCCAG
TGACTGGGGATAGTGAAGGTGCTACGGTGCAGGTAAAGTTCAGTGAGCTGCTCTGGGGAG
5 GGAAGGGACATAGAAGACTGTTCCATCATTCATTGCTTTTAAGGATGAGTTCTCTCTTGT
CAAATGCACTTCTGCCAGCAGACACCAGTTAAGTGGCGTTCATGGGGGTCTTTTCGCTGC
AGCCTCCACCGTGCTGAGGTGAGGAGGCCGACGTGGCAGTTGTGGTCCCTTTTGTCTGTA
TTAATGGCTGCTGACCTTCCAAAGCACTTTTTATTTTCATTTTCTGTACAGACACTCAG
GGATAGCAGTACCATTTTACTTCCGCAAGCCTTTAACTGCAAGATGAAGCTGCAAAGGGT
10 TTGAAATGGGAAGGTTTGAGTTCAGGCAGCGTATGAACTCTGGAGAGGGGCTGCCAGTC
CTCTCTGGGCCGCGAGCGGACCCAGCTGGAACACAGGAAGTTGGAGCAGTAGGTGCTCCTT
CACCTCTCAGTATGTCTCTTTCAACTCTAGTTTTTGAAGTGGGGACACAGGAAGTCCAGT
GGGGACACAGCCACTCCCCAAAGAATAAGGAACTTCCATGCTTCATTCCCTGGCATAAAA
AGTGNTCAAACACACCAGAGGGGGCAGGCACCAGCCAGGGTATGATGGGTACTACCCTTT
15 TCTGGAGAACCATAGACTTCCCTTACTACAGGGACTTGCATGTCCTAAAGCACTGGCTGA
AGGAAGCCAAGAGGATCACTGCTGCTCCTTTTTTGTAGAGGAAATGTTTGTGTACGTGGT
AAGATATGACCTAGCCCTTTTAGGTAAGCGAACTGGTATGTTAGTAACGTGTACAAAGTT
TAGGTTTACAGACCCCGGAGTCTTGGGCATGTGGGTCTCGGGTCACTGGTTTTGACTTTAG
GGCTTTGTTACAGATGTGTGACCAAGGGGAAAATGTGCATGACAACACTAGAGGTAGGGG
20 CGAAGCCAGAAAGAAGGGAAGTTTTGGCTGAAGTAGGAGTCTTGGTGAGATTTTGCTGTG
ATGCATGGTGTGAACTTTCTGAGCCTCTTGTTTTTCTCAGCTGACTCCATATTTTCCTA
CTTGTGGCAGCGACTGCATCCGACATAAAGGAACAGTTGTGCTCTGCCCACAAACAGGCG
TCCCTTTTCCCTCTGGATAACAACAAAAGCAAGCCGGGAGGCTGGCTGCCTCTCCTCCTGC
TGTCTCTGCTGGTGGCCACATGGGTGCTGGTGGCAGGGATCTATCTAATGTGGAGGCACG
25 AAAGGATCAAGAAGACTTCCTTTTCTACCACCACACTACTGCCCCCATTAAGGTTCTTG
TGGTTTACCCATCTGAAATATGTTTCCATCACACAATTTGTTACTTCACTGAATTTCTTC
AAAACCATTTGCAGAAGTGAGGTCATCCTTGAAAAGTGGCAGAAAAAGAAAATAGCAGAGA
TGGGTCCAGTGCAGTGGCTTGCCACTCAAAGAAGGCAGCAGACAAAGTCGTCTTCCTTC
TTTCCAATGACGTCAACAGTGTGTGCGATGGTACCTGTGGCAAGAGCGAGGGCAGTCCCA
30 GTGAGAACTCTCAAGACCTCTTCCCCCTTGCCCTTTAACCTTTTCTGCAGTGATCTAAGAA
GCCAGATTCATCTGCACAAATACGTGGTGGTCTACTTTAGAGAGATTGATACAAAAGACG
ATTACAATGCTCTCAGTGTCTGCCCCAAGTACCACTTCATGAAGGATGCCACTGCTTTCT
GTGCAGAACTTCTCCATGTCAAGCAGCAGGTGTCAGCAGGAAAAAGATCACAAGCCTGCC
ACGATGGCTGCTGCTCCTTGTAGCCCAACCATGAGAAGCAAGAGACCTTAAAGGCTTCCT
35 ATCCCACCAATTACAGGGAAAAAACGTGTGATGATCCTGAAGCTTACTATGCAGCCTACA
AACAGCCTTAGTAATTAACCATTTTATACCAATAAAATTTTCAAATATTACTAACTAAT
GTAGCATTAACCTAACGATTGGAACTACATTTACAACCTTCAAAGCTGTTTTATACATAGA
AATCAATTACAGCTTTAATTGAAAACCTGTAACCATTTTGATAATGCAACAATAAAGCATC
TTCCAAAAA

40

In another set of preferred embodiments of the invention, any sequence, or unique portion thereof, of the CACNA1D sequences of the I.M.A.G.E. Consortium cluster NM_000720, as well as

the UniGene Homo sapiens cluster Hs.399966, may be used. Similarly, any sequence encoding all or a part of the protein encoded by any CACNA1D sequence disclosed herein may be used. The consensus sequence of the I.M.A.G.E. Consortium cluster is as follows, with the assigned coding region (ending with a termination codon) underlined and preceded by the 5' untranslated and/or non-coding region and followed by the 3' untranslated and/or non-coding region:

SEQ ID NO:4 (consensus sequence for CACNA1D, identified as NM_000720 or NM_000720.1)

```
10  agaataaggg cagggaccgc ggctcctatc tcttggtgat ccccttcccc attccgcccc
    cgctcaacg cccagcacag tgccctgcac acagtagtcg ctcaataaat gttcgtggat
    gatgatgatg atgatgatga aaaaaatgca gcatcaacgg cagcagcaag cggaccacgc
    gaacgaggca aactatgcaa gaggcaccag acttcctctt tctggtgaag gaccaacttc
    tcagccgaat agctccaagc aaactgtcct gtcttggaag gctgcaatcg atgctgctag
    acaggccaag gctgccccaa ctatgagcac ctctgcaccc ccacctgtag gatctctctc
15  ccaaagaaaa cgctagcaat acgccaagag caaaaaacag ggtaactcgt ccaacagccg
    acctgcccgc gcccttttct gtttatcact caataacccc atccgaagag cctgcattag
    tatagtggaa tggaaaccat ttgacatatt tatattattg gctatttttg ccaattgtgt
    ggccttagct atttacatcc cattccctga agatgattct aattcaacaa atcataactt
    ggaaaaagta gaatatgcct tcctgattat ttttacagtc gagacatttt tgaagattat
20  agcgtatgga ttattgctac atcctaattgc ttatgttagg aatggatgga atttactgga
    ttttggtata gtaatagtag gattgtttag tgtaattttg gaacaattaa ccaaagaaac
    agaaggcggg aaccactcaa gcggcaaatc tggaggcttt gatgtcaaag ccctccgtgc
    ctttcgagtg ttgcgaccac ttcgactagt gtcaggggtg cccagtttac aagttgtcct
    gaactccatt ataaaagcca tggttccctt ccttcacata gcccttttgg tattattttg
25  aatcataatc tatgctatta taggattgga actttttatt ggaaaaatgc acaaaacatg
    tttttttgct gactcagata tcgtagctga agaggacca gctccatgtg cgttctcagg
    gaatggacgc cagtgtactg ccaatggcac ggaatgtagg agtggctggg ttggcccgaa
    cggaggcatc accaactttg ataactttgc ctttgccatg cttactgtgt ttcagtgcac
    caccatggag ggctggacag acgtgctcta ctgggtaaag gatgcatag gatgggaatg
30  gccatgggtg tattttgtta gtctgatcat ccttggctca ttttctgtcc ttaacctggg
    tcttggtgtc cttagtggag aattctcaaa ggaaagagag aaggcaaaag cacggggaga
    tttccagaag ctccgggaga agcagcagct ggaggaggat ctaaagggct acttggattg
    gatcacccaa gctgaggaca tcgatccgga gaatgaggaa gaaggaggag aggaaggcaa
    acgaaatact agcatgccc cagcgcagac tgagtctgtg aacacagaga acgtcagcgg
35  tgaaggcgag aaccgaggct gctgtggaag tctctggtgc tgggtggagac ggagaggcgc
    ggccaaggcg gggccctctg ggtgtcggcg gtgggggtcaa gccatctcaa aatccaaact
    cagccgacgc tggcgtcgct ggaaccgatt caatcgaga agatgtaggg ccgccgtgaa
    gtctgtcacg ttttactggc tggttatcgt cctggtgttt ctgaacacct taaccatttc
    ctctgagcac tacaatcagc cagattggtt gacacagatt caagatattg ccaacaaagt
40  cctcttggct ctgttcacct gcgagatgct ggtaaaaatg tacagcttgg gcctccaagc
    atatttcgtc tctcttttca accggtttga ttgcttcgtg gtgtgtggtg gaatcactga
```


gacgatcctg gtggaactgg aaatcatgtc tcccctgggg atctctgtgt ttcggtgtgt
gcgctcttta agaatcttca aagtgaccag gcactggact tccctgagca acttagtggc
atccttatta aactccatga agtccatcgc ttcgctgttg cttctgcttt ttctcttcat
tatcatcttt tcttgcttg ggatgcagct gtttggcggc aagttaatt ttgatgaaac
5 gcaaaccaag cggagcacct ttgacaattt cctcaagca cttctcacag tgttcagat
cctgacaggc gaagactgga atgctgtgat gtacgatggc atcatggctt acggggggcc
atcctcttca ggaatgatcg tctgcatcta cttcatcadc ctcttcattt gtggtaacta
tattctactg aatgtcttct tggccatcgc tgtagacaat ttggctgatg ctgaaagtct
gaacactgct cagaaagaag aagcggaga aaaggagagg aaaaagattg ccagaaaaga
10 gagcctagaa aataaaaaga acaacaaacc agaagtcaac cagatagcca acagtgacaa
caaggttaca attgatgact atagagaaga ggatgaagac aaggaccctt atccgccttg
cgatgtgcca gtaggggaag aggaagagga agaggaggag gatgaacctg aggttcttgc
cggaccccg cctcgaagga tctcggagtt gaacatgaag gaaaaaattg ccccatccc
tgaagggagc gctttcttca ttcttagcaa gaccaaccg atccgcgtag gctgccacaa
15 gctcatcaac caccacatct tcaccaacct catccttgtc ttcacatgc tgagcagcgc
tgccctggcc gcagaggacc ccatccgcag ccactcctc cggaacacga tactgggtta
ctttgactat gccttcacag ccatctttac tgttgagatc ctgttgaaga tgacaacttt
tggagctttc ctccacaaag gggccttctg caggaactac ttcaatttgc tggatatgct
gggtggttggg gtgtctcttg tgtcatttgg gattcaatcc agtgccatct ccgttgtgaa
20 gattctgagg gtcttaaggg tctgctgccc cctcagggcc atcaacagag caaaaggact
taagcacgtg gtccagtgcg tcttcgtggc catccggacc atcggcaaca tcatgatcgt
cactaccctc ctgcagttca tgtttgcctg tatcggggtc cagttgttca aggggaagt
ctatcgctgt acggatgaag ccaaaagtaa cctgaagaa tgcaggggac ttttcatcct
ctacaaggat ggggatgttg acagtctgt ggtccgtgaa cggatctggc aaaacagtga
25 tttcaacttc gacaacgtcc tctctgctat gatggcgctc ttcacagtct ccacgtttga
gggctggcct gcgttgctgt ataaagccat cgactcgaat ggagagaaca tgggccaat
ctacaaccac cgcgtggaga tctccatctt cttcatcadc tacatcatca ttgtagcttt
cttcatgatg aacatctttg tgggctttgt catcggttaca tttcaggaac aaggagaaaa
agagtataag aactgtgagc tggacaaaaa tcagcgtcag tgtgttgaat acgccttgaa
30 agcacgtccc ttgcggagat acatcccaa aaaccctac cagtacaagt tctggtacgt
ggtgaactct tcgcctttcg aatacatgat gtttgtctc atcatgctca acacactctg
cttggccatg cagcactacg agcagttcaa gatgttcaat gatgccatgg acattctgaa
catggtcttc accggggtgt tcaccgtcga gatgggtttg aaagtcacg catttaagcc
taaggggtat tttagtgaag cctggaacac gtttgactcc ctcatcgtaa tgggcagcat
35 tatagacgtg gccctcagcg aagcggacc aactgaaagt gaaaatgtcc ctgtcccaac
tgctacacct gggaactctg aagagagcaa tagaatctcc atcacctttt tccgtctttt
ccgagtgatg cgattggtga agcttctcag caggggggaa ggcacccgga cattgctgtg
gacttttatt aagtccttcc aggcgctccc gtatgtggcc ctctcatag ccatgctgtt
cttcatctat gcggtcattg gcatgcagat gtttgggaaa gttgccatga gagataacaa
40 ccagatcaat aggaacaata acttccagac gtttcccag gcggtgctgc tgctcttcag
gtgtgcaaca ggtgaggcct ggcaggagat catgctggcc tgtctcccag ggaagctctg
tgacctgag tcagattaca accccgggga ggagtataca tgtgggagca actttgccat
tgtctatttc atcagttttt acatgctctg tgcatttctg atcatcaatc tgtttgtggc
tgtcatcatg gataatttct actatctgac ccgggactgg tctatttttg ggcctcacca
45 tttagatgaa ttcaaaagaa tatggtcaga atatgaccct gaggcaaagg gaaggataaa

acaccttgat gtggtcactc tgcttcgacg catccagcct cccctgggggt ttgggaagtt
atgtccacac agggtagcgt gcaagagatt agttgccatg aacatgcctc tcaacagtga
cgggacagtc atgtttaatg caaccctggt tgctttgggt cgaacggctc ttaagatcaa
gaccgaaggg aacctggagc aagctaataga agaacttcgg gctgtgataa agaaaatttg
5 gaagaaaacc agcatgaaat tacttgacca agttgtccct ccagctgggtg atgatgaggt
aaccgtgggg aagttctatg ccactttcct gatacaggac tacttttagga aattcaagaa
acggaaagaa caaggactgg tgggaaagta ccctgcgaag aacaccacaa ttgccctaca
ggcgggatta aggacactgc atgacattgg gccagaaatc cggcgtgcta tatcgtgtga
tttgcaagat gacgagcctg aggaacaaaa acgagaagaa gaagatgatg tgttcaaaag
10 aaatgggtgcc ctgcttgga accatgtcaa tcatgttaat agtgatagga gagattccct
tcagcagacc aataccaccc accgtccctt gcatgtccaa aggccttcaa ttccacctgc
aagtgatact gagaaaccgc tgtttcctcc agcaggaaat tcggtgtgtc ataaccatca
taaccataat tccataggaa agcaagttcc cacctcaaca aatgccaatc tcaataatgc
caatatgtcc aaagctgccc atggaaagcg gcccagcatt gggaaccttg agcatgtgtc
15 tgaaaatggg catcattctt cccacaagca tgaccgggag cctcagagaa ggtccagtgt
gaaaagaacc cgctattatg aaacttacat taggtccgac tcaggagatg aacagctccc
aactatttgc cgggaagacc cagagataca tggctatttc agggaccccc actgcttggg
ggagcaggag tatttcagta gtgaggaatg ctacgaggat gacagctcgc ccacctggag
caggcaaaac tatggctact acagcagata cccaggcaga aacatcgact ctgagaggcc
20 ccgaggctac catcatcccc aaggattctt ggaggacgat gactcgcccc tttgctatga
ttcacggaga tctccaagga gacgcctact acctcccacc ccagcatccc accggagatc
ctccttcaac tttgagtgcc tgcgccggca gagcagccag gaagagggtcc cgtcgtctcc
catcttcccc catcgcacgg ccctgcctct gcatctaatag cagcaacaga tcatggcagt
tgccggccta gattcaagta aagcccagaa gtactcaccg agtcaactga cccggctcgtg
25 ggccacccct ccagcaaccc ctccctaccg ggactggaca ccgtgctaca cccccctgat
ccaagtggag cagtcagagg ccctggacca ggtgaacggc agcctgccgt ccctgcaccg
cagctcctgg tacacagacg agcccagacat ctctaccgg actttcacac cagccagcct
gactgtcccc agcagcttcc ggaacaaaaa cagcgacaag cagaggagtg cggacagctt
gggtggaggca gtccctgatat ccgaaggctt gggacgctat gcaagggacc caaaatttgt
30 gtcagcaaca aaacacgaaa tcgctgatgc ctgtgacctc accatcgacg agatggagag
tgcagccagc accctgctta atgggaacgt gcgtccccga gccaacgggg atgtgggccc
cctctcacac cggcaggact atgagctaca ggactttgggt cctggctaca gcgacgaaga
gccagaccct gggaggggatg aggaggacct ggccgatgaa atgatatgca tcaccacctt
gtagccccca gcgaggggca gactggctct ggccctcaggt ggggcgcagg agagccaggg
35 gaaaagtgcc tcatagttag gaaagttag gactagttag ggagtaatat tcaattaatt
agacttttgt ataagagatg tcatgcctca agaaagccat aaacctggta ggaacaggtc
ccaagcgggt gagcctggca gagtaccatg cgctcgcccc cagctgcagg aaacagcagg
ccccgcctc tcacagagga tgggtgagga ggccagacct gccctgcccc attgtccaga
tgggcaactgc tgtggagtct gcttctccca tgtaccaggg caccaggccc acccaactga
40 aggcattggcg gcgggggtgca ggggaaagt aaaggtgatg acgatcatca cacctcgtgt
cgttacctca gccatcggtc tagcatatca gtactgggc ccaacatatc catttttaaa
ccctttcccc caaatacact gcgtcctgggt tcctgttttag ctgttctgaa ata

I.M.A.G.E. Consortium Clone ID numbers and the corresponding GenBank accession numbers of sequences identified as belonging to the I.M.A.G.E. Consortium and UniGene clusters, are listed below. Also included are sequences that are not identified as having a Clone ID number but still identified as being those of CACNA1D. The sequences include those of the "sense" and complementary strands sequences corresponding to CACNA1D. The sequence of each GenBank accession number is presented in the attached Appendix.

Clone ID numbers	GenBank accession numbers
5676430	BM128550
5197948	BI755471
6027638	BQ549084, BQ549571
2338956	AI693324
36581	R25307, R46658
49630	H29256, H29339
4798765	BG716371
2187310	AI537488
838231	AA458692
2111614	AI393327
2183482	AI520947
1851007	AI248998
1675503	AI075844
2434923	AI869807
2434924	AI869800
1845827	AI243110
2511756	AI955764
628568	AA192669, AA192157
2019331	AI361691
2337381	AI914244
2503579	AW008769
2503626	AW008794
1160989	AA877582
1653475	AI051972

1627755	AI017959
287750	N79331, N62240
1867677	AI240933
1618303	AI015031
1881344	AI290994
1408031	AA861160
1557035	AA915941
956303	AA493341
2148234	AI467998
1499899	AA885585
1647592	AI033648
2341185	AI697633
981603	AA523647
6281678	BQ710377
6278348	BQ706920
5876024	BQ016847
6608849	CA943595
5440464	BM008196
5209489	BI769856
5183025	BI758971
880540	AA468565
757337	AA437099
6608849	CA867864
461797	AA682690
434787	AA701888
6151588	BU182632
6295618	BQ898429
6300779	BQ711800
434811	AA703120
1568025	AA978315
3220210	BE550599
3214121	BE502741
3009312	AW872382

2733394	AW444663
2872156	AW341279
30514550	CF456750
2718456	AW139850
2543682	AW029633
2492730	AI963788
2545866	AI951788
2272081	AI680744
2152336	AI601252
2146429	AI459166
1274498	AA885750
2272081	BX092736
287750	BX114568
3233645	BE672659
289209	N78509, N73668
277086	N46744, N39597
3272340	BF439267
3273859	BF436153
3568401	BF110611
None (mRNA sequences)	M76558, AF088004, M83566
None	CB410657, BQ372430, BQ366601, BQ324528, BQ318830, AL708030, BM509161, N85902, BQ774355, CA774243, CA436347, CA389011, BU679327, BU608029, BU073743, BE175413, AW969248, AI908115, BF754485, BI015409, BG202552, BF883669, BF817590, BF807128, BF806160, BF805244, BF805235, BF805080, T27949, BE836638, BE770685, BE769065,

In one preferred embodiment, any sequence, or unique portion thereof, of the following CACNA1D sequence, identified by AF088004 or AF088004.1, may be used in the practice of the invention.

5

SEQ ID NO:5 (sequence for CACNA1D):

TTTTTTTTTTTTTTTTTTTTTTCTTACAAAGAAAAATTTAATATTCGATGAGAGGTTGAAC
CAGGCTTAAAGCAGACATACTAGGAAATGGTGCAGCCTGTAAGAATGCCAGTTTGTAAGT
ACTGACTTTGGAAAAGATCATCGCCTCTATCAGACACTTAGGGTCCTGGTCTGGCAATTT
5 TGGCCTGATGTGATGCCACAAGACCCAACAGAGAGAGACACAGAGTCCAGGATAATGTTG
ACAGTGGTGTAGCCCTTTAGGAGAAATGGCGCTCCCTGCGGCTGGTATTAGGTTACCATT
GGCACCGAAGGAACCAGGAGGATAAGAATATCCATAATTTAGAGCTGCCCTGGCACAGT
ACCTGCCCCGTCGGAGGCTCTCACTGGCAAATGACAGCTCTGTGCAAGGAGCACTCCCAA
GTATAAAATTTATTACACAGTTTTATTCTGAAGAACATTTTGCATTTTAATAAAAAAGGA
10 TTTATGTCAGGAAAGAGTCATTTACAAACCTTGAAGTGTTTTTGCCTGGATCAGAGTAAG
AATGTCTTAAGAAGAGGTTTGTAAAGTCTTCATAACAAAGTGGTGTGTGTTATTTACAAA
AAAAAAAAAAAAAAAAAATTAACAGGTTGTCTGTATACTATTAAAAATTTTGGACCAAAAA
AAAAAAAAAAAAAAAAA

15 In another set of preferred embodiments of the invention, any sequence, or unique portion thereof, of the HOXB13 sequences of the I.M.A.G.E. Consortium cluster NM_006361, as well as the UniGene Homo sapiens cluster Hs.66731, may be used. Similarly, any sequence encoding all or a part of the protein encoded by any HOXB13 sequence disclosed herein may be used. The consensus sequence of the I.M.A.G.E. Consortium cluster is as follows, with the assigned coding
20 region (ending with a termination codon) underlined and preceded by the 5' untranslated and/or non-coding region and followed by the 3' untranslated and/or non-coding region:

SEQ ID NO:6 (consensus sequence for HOXB13, identified as NM_006361 or NM_006361.2)

25 cgaatgcagg cgacttgcca gctgggagcg atttaaaacg ctttggattc ccccggcctg
ggtggggaga gcgagctggg tgccccctag attccccgcc ccgcacctc atgagccgac
cctcggctcc atggagcccg gcaattatgc caccttggat ggagccaagg atatcgaagg
cttgctggga gcgggagggg ggcggaatct ggtcgccac tccctctga ccagccaccc
agcggcgct acgctgatgc ctgctgtcaa ctatgcccc ttggatctgc caggctcggc
30 ggagccgcca aagcaatgcc acccatgccc tggggtgccc caggggacgt cccagctcc
cgtgccttat ggttactttg gaggcgggta ctactcctgc cgagtgtccc ggagctcgt
gaaaccctgt gccaggcag ccaccctggc cgcgtacccc gcggagactc ccacggccgg
ggaagagtac ccagtcgcc cactgagtt tgccttctat cgggatatac cgggaaccta
ccacgctatg gccagttacc tggacgtgtc tgtggtgcag actctgggtg ctctggaga
35 accgcgacat gactccctgt tgctgtgga cagttaccag tcttgggctc tcgctggtgg
ctggaacagc cagatgtgtt gccagggaga acagaacca ccaggccct tttggaaggc
agcatttgca gactccagcg ggcagcacc tctgacgcc tgcgccttgc gtcgcgccg
caagaaacgc attccgtaca gcaaggggca gttgcgggag ctggagcggg agtatgcggc
taacaagttc atcaccaagg acaagaggcg caagatctcg gcagccacca gcctctcgg

5 gcgccagatt accatctggt ttcagaaccg ccgggtcaaa gagaagaagg ttctcgccaa
ggtgaagaac agcgctaccc ctttaagagat ctccttgccct ggggtgggagg agcgaaagtg
ggggtgtcct ggggagacca gaaacctgcc aagcccaggc tggggccaag gactctgctg
agaggccct agagacaaca cccttcccag gccactggct gctggactgt tcctcaggag
cggcctgggt acccagtatg tgcagggaga cggaaccca tgtgacaggc ccactccacc
agggttcca aagaacctgg ccagtcata atcattcatt ctcacagtgg caataatcac
gataaccagt

10 I.M.A.G.E. Consortium Clone ID numbers and the corresponding GenBank accession numbers of sequences identified as belonging to the I.M.A.G.E. Consortium and UniGene clusters, are listed below. Also included are sequences that are not identified as having a Clone ID number but still identified as being those of HOXB13. The sequences include those of the "sense" and complementary strands sequences corresponding to HOXB13. The sequence of each GenBank accession number is presented in the attached Appendix.

15

Clone ID numbers	GenBank accession numbers
4250486	BF676461, BC007092
5518335	BM462617
4874541	BG752489
4806039	BG778198
3272315	CB050884, CB050885
4356740	BF965191
6668163	BU930208
1218366	AA807966
2437746	AI884491
1187697	AA652388
3647557	BF446158
1207949	AA657924
1047774	AA644637
3649397	BF222357
971664	AA527613
996191	AA533227
813481	AA456069, AA455572, BX117624
6256333	BQ673782

2408470	AI814453
2114743	AI417272
998548	AA535663
2116027	AI400493
3040843	AW779219
1101311	AA594847
1752062	AI150430
898712	AA494387
1218874	AA662643
2460189	AI935940
986283	AA532530
1435135	AA857572
1871750	AI261980
3915135	BE888751
2069668	AI378797
667188	AA234220, AA236353
1101561	AA588193
1170268	AI821103, AI821851, AA635855
2095067	AI420753
4432770	BG180547
783296	AA468306, AA468232
3271646	CB050115, CB050116
1219276	AA661819
30570598	CF146837
30570517	CF146763
30568921	CF144902
3099071	CF141511
3096992	CF139563
3096870	CF139372
3096623	CF139319
3096798	CF139275
30572408	CF122893
2490082	AI972423

2251055	AI918975
2419308	AI826991
2249105	AI686312
2243362	AI655923
30570697	CF146922
3255712	BF476369
3478356	BF057410
3287977	BE645544
3287746	BE645408
3621499	BE388501
30571128	CF147366
30570954	CF147143
None (mRNA sequences)	BT007410, BC007092, U57052, U81599
None	CB120119, CB125764, AU098628, CB126130, BI023924, BM767063, BM794275, BQ363211, BM932052, AA357646, AW609525, CB126919, AW609336, AW609244, BF855145, AU126914, CB126449, AW582404, BX641644

In one preferred embodiment, any sequence, or unique portion thereof, of the following HOXB13 sequence, identified by BC007092 or BC007092.1, may be used in the practice of the invention.

5

SEQ ID NO:7 (sequence for HOXB13):

10 GGATTCCCCCGGCCTGGGTGGGGAGAGCGAGCTGGGTGCCCCCTAGATTCCCCGCCCCCG
CACCTCATGAGCCGACCCTCGGCTCCATGGAGCCCGGCAATTATGCCACCTTGGATGGAG
CCAAGGATATCGAAGGCTTGCTGGGAGCGGGAGGGGGCGGAATCTGGTCGCCCCACTCCC
CTCTGACCAGCCACCCAGCGGCGCCTACGCTGATGCCTGCTGTCAACTATGCCCCCTTGG
ATCTGCCAGGCTCGGCGGAGCCGCCAAAGCAATGCCACCCATGCCCTGGGGTGGCCCCAGG
GGACGTCCCCAGCTCCCGTGCCTTATGGTTACTTTGGAGGCGGGTACTACTCCTGCCGAG
15 TGTCCCGGAGCTCGCTGAAACCCTGTGCCAGGCAGCCACCCTGGCCGCGTACCCCGCGG
AGACTCCACGGCCGGGAAGAGTACCCAGCCGCCCCACTGAGTTTGCCTTCTATCCGG
GATATCCGGGAACCTACCAGCCTATGGCCAGTTACCTGGACGTGTCTGTGGTGCAGACTC
TGGGTGCTCCTGGAGAACCGCGACATGACTCCCTGTTGCCTGTGGACAGTTACCACTCTT
GGGCTCTCGCTGGTGGCTGGAACAGCCAGATGTGTTGCCAGGGAGAACAGAACCCACCAG

GTCCCTTTTGGAAAGGCAGCATTTGCAGACTCCAGCGGGCAGCACCCCTCCTGACGCCTGCG
CCTTTCGTCGCGGCCGCAAGAAACGCATTCCGTACAGCAAGGGGCAGTTGCGGGAGCTGG
AGCGGGAGTATGCGGCTAACAAGTTCATCACCAAGGACAAGAGGCGCAAGATCTCGGCAG
CCACCAGCCTCTCGGAGCGCCAGATTACCATCTGGTTTCAGAACCGCCGGGTCAAAGAGA
5 AGAAGGTTCTCGCCAAGGTGAAGAACAGCGCTACCCCTTAAGAGATCTCCTTGCCTGGGT
GGGAGGAGCGAAAGTGGGGGTGTCCTGGGGAGACCAGGAACCTGCCAAGCCCAGGCTGGG
GCCAAGGACTCTGCTGAGAGGCCCTTAGAGACAACACCCTTCCCAGGCCACTGGCTGCTG
GACTGTTCTCAGGAGCGGCCTGGGTACCCAGTATGTGCAGGGAGACGGAACCCCATGTG
ACAGCCCACTCCACCAGGGTTCCTCAAAGAACCTGGCCCAAGTCATAATCATTCATCCTGAC
10 AGTGGCAATAATCACGATAACCAGTACTAGCTGCCATGATCGTTAGCCTCATATTTTCTA
TCTAGAGCTCTGTAGAGCACTTTAGAAACCGCTTTCATGAATTGAGCTAATTATGAATAA
ATTTGGAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

Sequences identified by SEQ ID NO. are provided using conventional representations of a
15 DNA strand starting from the 5' phosphate linked end to the 3' hydroxyl linked end. The
assignment of coding regions is generally by comparison to available consensus sequence(s) and
therefore may contain inconsistencies relative to other sequences assigned to the same cluster.
These have no effect on the practice of the invention because the invention can be practiced by use
of shorter segments (or combinations thereof) of sequences unique to each of the three sets
20 described above and not affected by inconsistencies. As non-limiting examples, a segment of
IL17BR, CACNA1D, or HOXB13 nucleic acid sequence composed of a 3' untranslated region
sequence and/or a sequence from the 3' end of the coding region may be used as a probe for the
detection of IL17BR, CACNA1D, or HOXB13 expression, respectively, without being affected by
the presence of any inconsistency in the coding regions due to differences between sequences.
25 Similarly, the use of an antibody which specifically recognizes IL17BR, CACNA1D, or HOXB13
protein to detect its expression would not be affected by the presence of any inconsistency in the
representation of the coding regions provided above.

As will be appreciated by those skilled in the art, some of the above sequences include 3'
poly A (or poly T on the complementary strand) stretches that do not contribute to the uniqueness of
30 the disclosed sequences. The invention may thus be practiced with sequences lacking the 3' poly A
(or poly T) stretches. The uniqueness of the disclosed sequences refers to the portions or entireties
of the sequences which are found only in IL17BR, CACNA1D, or HOXB13 nucleic acids,
including unique sequences found at the 3' untranslated portion of the genes. Preferred unique
sequences for the practice of the invention are those which contribute to the consensus sequences

for each of the three sets such that the unique sequences will be useful in detecting expression in a variety of individuals rather than being specific for a polymorphism present in some individuals. Alternatively, sequences unique to an individual or a subpopulation may be used. The preferred unique sequences are preferably of the lengths of polynucleotides of the invention as discussed
5 herein.

To determine the (increased or decreased) expression levels of the above described sequences in the practice of the present invention, any method known in the art may be utilized. In one preferred embodiment of the invention, expression based on detection of RNA which hybridizes to polynucleotides containing the above described sequences is used. This is readily
10 performed by any RNA detection or amplification+detection method known or recognized as equivalent in the art such as, but not limited to, reverse transcription-PCR (optionally real-time PCR), the methods disclosed in U.S. Patent Application 10/062,857 entitled "Nucleic Acid Amplification" filed on October 25, 2001 as well as U.S. Provisional Patent Applications
15 60/298,847 (filed June 15, 2001) and 60/257,801 (filed December 22, 2000), the methods disclosed in U.S. Patent 6,291,170, and quantitative PCR. Methods to identify increased RNA stability (resulting in an observation of increased expression) or decreased RNA stability (resulting in an observation of decreased expression) may also be used. These methods include the detection of sequences that increase or decrease the stability of mRNAs containing the IL17BR, CACNA1D, or HOXB13 sequences disclosed herein. These methods also include the detection of increased
20 mRNA degradation.

In particularly preferred embodiments of the invention, polynucleotides having sequences present in the 3' untranslated and/or non-coding regions of the above disclosed sequences are used to detect expression or non-expression of IL17BR, CACNA1D, or HOXB13 sequences in breast cells in the practice of the invention. Such polynucleotides may optionally contain sequences found
25 in the 3' portions of the coding regions of the above disclosed sequences. Polynucleotides containing a combination of sequences from the coding and 3' non-coding regions preferably have the sequences arranged contiguously, with no intervening heterologous sequence(s).

Alternatively, the invention may be practiced with polynucleotides having sequences present in the 5' untranslated and/or non-coding regions of IL17BR, CACNA1D, or HOXB13 sequences in

breast cells to detect their levels of expression. Such polynucleotides may optionally contain sequences found in the 5' portions of the coding regions. Polynucleotides containing a combination of sequences from the coding and 5' non-coding regions preferably have the sequences arranged contiguously, with no intervening heterologous sequence(s). The invention may also be practiced with sequences present in the coding regions of IL17BR, CACNA1D, or HOXB13.

Preferred polynucleotides contain sequences from 3' or 5' untranslated and/or non-coding regions of at least about 16, at least about 18, at least about 20, at least about 22, at least about 24, at least about 26, at least about 28, at least about 30, at least about 32, at least about 34, at least about 36, at least about 38, at least about 40, at least about 42, at least about 44, or at least about 46 consecutive nucleotides. The term "about" as used in the previous sentence refers to an increase or decrease of 1 from the stated numerical value. Even more preferred are polynucleotides containing sequences of at least or about 50, at least or about 100, at least or about 150, at least or about 200, at least or about 250, at least or about 300, at least or about 350, or at least or about 400 consecutive nucleotides. The term "about" as used in the preceding sentence refers to an increase or decrease of 10% from the stated numerical value.

Sequences from the 3' or 5' end of the above described coding regions as found in polynucleotides of the invention are of the same lengths as those described above, except that they would naturally be limited by the length of the coding region. The 3' end of a coding region may include sequences up to the 3' half of the coding region. Conversely, the 5' end of a coding region may include sequences up the 5' half of the coding region. Of course the above described sequences, or the coding regions and polynucleotides containing portions thereof, may be used in their entirety.

Polynucleotides combining the sequences from a 3' untranslated and/or non-coding region and the associated 3' end of the coding region are preferably at least or about 100, at least or about 150, at least or about 200, at least or about 250, at least or about 300, at least or about 350, or at least or about 400 consecutive nucleotides. Preferably, the polynucleotides used are from the 3' end of the gene, such as within about 350, about 300, about 250, about 200, about 150, about 100, or about 50 nucleotides from the polyadenylation signal or polyadenylation site of a gene or expressed sequence. Polynucleotides containing mutations relative to the sequences of the disclosed genes

may also be used so long as the presence of the mutations still allows hybridization to produce a detectable signal.

In another embodiment of the invention, polynucleotides containing deletions of nucleotides from the 5' and/or 3' end of the above disclosed sequences may be used. The deletions are preferably of 1-5, 5-10, 10-15, 15-20, 20-25, 25-30, 30-35, 35-40, 40-45, 45-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-125, 125-150, 150-175, or 175-200 nucleotides from the 5' and/or 3' end, although the extent of the deletions would naturally be limited by the length of the disclosed sequences and the need to be able to use the polynucleotides for the detection of expression levels.

Other polynucleotides of the invention from the 3' end of the above disclosed sequences include those of primers and optional probes for quantitative PCR. Preferably, the primers and probes are those which amplify a region less than about 350, less than about 300, less than about 250, less than about 200, less than about 150, less than about 100, or less than about 50 nucleotides from the from the polyadenylation signal or polyadenylation site of a gene or expressed sequence.

In yet another embodiment of the invention, polynucleotides containing portions of the above disclosed sequences including the 3' end may be used in the practice of the invention. Such polynucleotides would contain at least or about 50, at least or about 100, at least about or 150, at least or about 200, at least or about 250, at least or about 300, at least or about 350, or at least or about 400 consecutive nucleotides from the 3' end of the disclosed sequences.

The invention thus also includes polynucleotides used to detect IL17BR, CACNA1D, or HOXB13 expression in breast cells. The polynucleotides may comprise a shorter polynucleotide consisting of sequences found in the above provided SEQ ID NOS in combination with heterologous sequences not naturally found in combination with IL17BR, CACNA1D, or HOXB13 sequences.

As non-limiting examples, a polynucleotide comprising one of the following sequences may be used in the practice of the invention.

SEQ ID NO:8:

CAATTACAGGGAAAAACGTGTGATGATCCTGAAGCTTACTATGCAGCCTACAAACAGCC

SEQ ID NO:9:

GCTCTCACTGGCAAATGACAGCTCTGTGCAAGGAGCACTCCCAAGTATAAAAATTATTAC

SEQ ID NO:10:

GATCGTTAGCCTCATATTTTCTATCTAGAGCTCTGTAGAGCACTTTAGAAACCGCTTTCA

5

Stated differently, the invention may be practiced with a polynucleotide consisting of the sequence of SEQ ID NOS:8, 9 or 10 in combination with one or more heterologous sequences that are not normally found with SEQ ID NOS:8, 9 or 10. Alternatively, the invention may also be practiced with a polynucleotide consisting of the sequence of SEQ ID NOS:8, 9 or 10 in combination with one or more naturally occurring sequences that are normally found with SEQ ID NOS:8, 9 or 10.

10

Polynucleotides with sequences comprising SEQ ID NOS:8 or 9, either naturally occurring or synthetic, may be used to detect nucleic acids which are over expressed in breast cancer cells that are responsive, and those which are not over expressed in breast cancer cells that are non-responsive, to treatment with TAM or another “antiestrogen” agent against breast cancer. Polynucleotides with sequences comprising SEQ ID NO:10, either naturally occurring or synthetic, may be used to detect nucleic acids which are under expressed in breast cancer cells that are responsive, and those which are not under expressed in breast cancer cells that are non-responsive, to treatment with TAM or another “antiestrogen” agent against breast cancer.

15

20

Additional sequences that may be used in polynucleotides as described above for SEQ ID NOS:8 and 9 are the following, wherein SEQ ID NOS:33 is complementary to a portion of IL17BR sequences disclosed herein:

SEQ ID NO:11:

25

TGCCTAATTTCACTCTCAGAGTGAGGCAGGTAAGTGGGGCTCCACTGGGTCCTCTGAGA

SEQ ID NO:12:

TTGGAAGCAGAGTCCCTCTAAAGGTAAGTCTTGTGGTCACTCAATATTGTATTGGCATT

30

SEQ ID NO:13:

ACGTTAGACTTTTGCTGGCATTCAAGTCATGGCTAGTCTGTGTATTTAATAAATGTGTGT

SEQ ID NO:14:

CTGGTCAGCCACTCTGACTTTTCTACCACATTAAATTCTCCATTACATCTCACTATTGGT

SEQ ID NO:15:

5 TACAACCTTCTGAATGCTGCACATTCTTCCAAAATGATCCTTAGCACAATCTATTGTATGA

SEQ ID NO:16:

GGGATGGCCTTTAGGCCACAGTAGTGTCTGTGTTAAGTTCATAAATGTGTATTTAATGA

SEQ ID NO:17:

10 CTCAAAGTGCTAAAGCTATGGTTGACTGCTCTGGTGTTTTTATATTCATTCGTGCTTTAG

SEQ ID NO:32:

15 CTGAAGCTTACTATGCAGCCTACAA

SEQ ID NO:33:

TCCAATCGTTAGTTAATGCTACATTAGTT

SEQ ID NO:34:

20 CAGCCTTAGTAATTAAAAC

Additional sequences that may be used in polynucleotides as described above for SEQ ID NO:10 are the following, wherein SEQ ID NOs:36 is complementary to a portion of IL17BR sequences disclosed herein:

SEQ ID NO:18:

25 CTATGGGGATGGTCCACTGTCACTGTTTCTCTGCTGTTGCAAATACATGGATAACACATT

SEQ ID NO:19:

30 ACTGGAAGAGCAGATGGTCTGACTGTGCTATGGCCTCATCATCAAGACTTTCAATCCTAT

SEQ ID NO:20:

ACGCCAAGCTCTTCAGTGAAGACACGATGTTATTAAAAGCCTGTTTTAGGGACTGCAAAA

SEQ ID NO:21:

35 TTTTGTAAAATCTTTAACCTTCCCTTTGTTCTTCATGTACACGCTGAACTGCAATTCTT

SEQ ID NO:22:

AACCTGGGGCATTTAGGGCAGAGGACAAAAGGATGTCAGCAATTGCTTGGGCTGCTTGGC

SEQ ID NO:23:

5 CTGGAACCTCTGGACTCCCCATGCTCTAACTCCCACACTCTGCTATCAGAACTTAACT

SEQ ID NO:24:

AACCCAGAACCATCTAAGACATGGGATTCAGTGATCATGTGGTTCTCCTTTTAACTTAC

SEQ ID NO:25:

10 GGCCATGTGCCATGGTATTTGGGTCCTGGGAGGGTGGGTGAAATAAAGGCATACTGTCTT

SEQ ID NO:26:

15 GTGTAGGCAGTCATGGCACCAAAGCCACCAGACTGACAAATGTGTATCAGATGCTTTTGT

SEQ ID NO:27:

GAAAACCTCTTCAAAAGACAAAAGCTGGCACTGCATTCTCTCTGTAGCAGGACAGAA

SEQ ID NO:28:

20 CACATCTTTAGGGTCAGTGAACAATGGGGCACATTTGGCACTAGCTTGAGCCCAACTCTG

SEQ ID NO:29:

GCCTTAATTTCTCATCTGAAAAGCTGGAAGGCCTGACTTGACTTGTTGAGCTTAAGATCC

SEQ ID NO:30:

25 CTTCAGGGGAGGATCAAGCTTTGAACCAAAGCCAATCACTGGCTTGATTTGTGTTTTTTA

SEQ ID NO:31:

30 ACAAGTTTTCACTGAATGAGCATGGCAGTGCCACTCAAGAAAATGAATCTCCAAAGTATC

SEQ ID NO:35:

GCCATGATCGTTAGCCTCATATT

SEQ ID NO:36:

35 CAATTCATGAAAGCGGTTTCTAAAG

SEQ ID NO:37:

TCTATCTAGAGCTCTGTAGAGC

Additionally, polynucleotides containing other sequences, particularly unique sequences, present in naturally occurring nucleic acid molecules comprising SEQ ID NOS:8-37 may be used in the practice of the invention.

Other polynucleotides for use in the practice of the invention include those that have
5 sufficient homology to those described above to detect expression by use of hybridization techniques. Such polynucleotides preferably have about or 95%, about or 96%, about or 97%, about or 98%, or about or 99% identity with IL17BR, CACNA1D, or HOXB13 sequences as described herein. Identity is determined using the BLAST algorithm, as described above. The other polynucleotides for use in the practice of the invention may also be described on the basis of
10 the ability to hybridize to polynucleotides of the invention under stringent conditions of about 30% v/v to about 50% formamide and from about 0.01M to about 0.15M salt for hybridization and from about 0.01M to about 0.15M salt for wash conditions at about 55 to about 65°C or higher, or conditions equivalent thereto.

In a further embodiment of the invention, a population of single stranded nucleic acid
15 molecules comprising one or both strands of a human IL17BR or CACNA1D sequence is provided as a probe such that at least a portion of said population may be hybridized to one or both strands of a nucleic acid molecule quantitatively amplified from RNA of a breast cancer cell. The population may be only the antisense strand of a human IL17BR or CACNA1D sequence such that a sense strand of a molecule from, or amplified from, a breast cancer cell may be hybridized to a portion of
20 said population. The population preferably comprises a sufficiently excess amount of said one or both strands of a human IL17BR or CACNA1D sequence in comparison to the amount of expressed (or amplified) nucleic acid molecules containing a complementary IL17BR or CACNA1D sequence from a normal breast cell. This condition of excess permits the increased amount of nucleic acid expression in a breast cancer cell to be readily detectable as an increase.

25 Alternatively, the population of single stranded molecules is equal to or in excess of all of one or both strands of the nucleic acid molecules amplified from a breast cancer cell such that the population is sufficient to hybridize to all of one or both strands. Preferred cells are those of a breast cancer patient that is ER+ or for whom treatment with tamoxifen or one or more other “antiestrogen” agent against breast cancer is contemplated. The single stranded molecules may of

course be the denatured form of any IL17BR and/or CACNA1D sequence containing double stranded nucleic acid molecule or polynucleotide as described herein.

The population may also be described as being hybridized to IL17BR or CACNA1D sequence containing nucleic acid molecules at a level of at least twice as much as that by nucleic acid molecules of a normal breast cell. As in the embodiments described above, the nucleic acid molecules may be those quantitatively amplified from a breast cancer cell such that they reflect the amount of expression in said cell.

The population is preferably immobilized on a solid support, optionally in the form of a location on a microarray. A portion of the population is preferably hybridized to nucleic acid molecules quantitatively amplified from a non-normal or abnormal breast cell by RNA amplification. The amplified RNA may be that derived from a breast cancer cell, as long as the amplification used was quantitative with respect to IL17BR or CACNA1D containing sequences.

In another embodiment of the invention, expression based on detection of DNA status may be used. Detection of the HOXB13 DNA as methylated, deleted or otherwise inactivated, may be used as an indication of decreased expression as found in non-normal breast cells. This may be readily performed by PCR based methods known in the art. The status of the promoter regions of HOXB13 may also be assayed as an indication of decreased expression of HOXB13 sequences. A non-limiting example is the methylation status of sequences found in the promoter region.

Conversely, detection of the DNA of a sequence as amplified may be used for as an indication of increased expression as found in non-normal breast cells. This may be readily performed by PCR based, fluorescent *in situ* hybridization (FISH) and chromosome *in situ* hybridization (CISH) methods known in the art.

A preferred embodiment using a nucleic acid based assay to determine expression is by immobilization of one or more of the sequences identified herein on a solid support, including, but not limited to, a solid substrate as an array or to beads or bead based technology as known in the art. Alternatively, solution based expression assays known in the art may also be used. The immobilized sequence(s) may be in the form of polynucleotides as described herein such that the polynucleotide would be capable of hybridizing to a DNA or RNA corresponding to the sequence(s).

The immobilized polynucleotide(s) may be used to determine the state of nucleic acid samples prepared from sample breast cancer cell(s), optionally as part of a method to detect ER status in said cell(s). Without limiting the invention, such a cell may be from a patient suspected of being afflicted with, or at risk of developing, breast cancer. The immobilized polynucleotide(s) need only be sufficient to specifically hybridize to the corresponding nucleic acid molecules derived from the sample (and to the exclusion of detectable or significant hybridization to other nucleic acid molecules).

In yet another embodiment of the invention, a ratio of the expression levels of two of the disclosed genes may be used to predict response to treatment with TAM or another SERM.

Preferably, the ratio is that of two genes with opposing patterns of expression, such as an underexpressed gene to an overexpressed gene, in correlation to the same phenotype. Non-limiting examples include the ratio of HOXB13 over IL17BR or the ratio of HOXB13 over CACNA1D. This aspect of the invention is based in part on the observation that such a ratio has a stronger correlation with TAM treatment outcome than the expression level of either gene alone. For example, the ratio of HOXB13 over IL17BR has an observed classification accuracy of 77%.

As a non-limiting example, the Ct values from Q-PCR based detection of gene expression levels may be used to derive a ratio to predict the response to treatment with one or more “antiestrogen” agent against breast cancer.

Additional Embodiments of the Invention

In embodiments where only one or a few genes are to be analyzed, the nucleic acid derived from the sample breast cancer cell(s) may be preferentially amplified by use of appropriate primers such that only the genes to be analyzed are amplified to reduce contaminating background signals from other genes expressed in the breast cell. Alternatively, and where multiple genes are to be analyzed or where very few cells (or one cell) is used, the nucleic acid from the sample may be globally amplified before hybridization to the immobilized polynucleotides. Of course RNA, or the cDNA counterpart thereof may be directly labeled and used, without amplification, by methods known in the art.

Sequence expression based on detection of a presence, increase, or decrease in protein levels or activity may also be used. Detection may be performed by any immunohistochemistry (IHC)

based, bodily fluid based (where a IL17BR, CACNA1D, and/or HOXB13 polypeptide is found in a bodily fluid, such as but not limited to blood), antibody (including autoantibodies against the protein where present) based, ex foliate cell (from the cancer) based, mass spectroscopy based, and image (including used of labeled ligand where available) based method known in the art and
5 recognized as appropriate for the detection of the protein. Antibody and image based methods are additionally useful for the localization of tumors after determination of cancer by use of cells obtained by a non-invasive procedure (such as ductal lavage or fine needle aspiration), where the source of the cancerous cells is not known. A labeled antibody or ligand may be used to localize the carcinoma(s) within a patient.

10 Antibodies for use in such methods of detection include polyclonal antibodies, optionally isolated from naturally occurring sources where available, and monoclonal antibodies, including those prepared by use of IL17BR, CACNA1D, and/or HOXB13 polypeptides as antigens. Such antibodies, as well as fragments thereof (including but not limited to F_{ab} fragments) function to detect or diagnose non-normal or cancerous breast cells by virtue of their ability to specifically bind
15 IL17BR, CACNA1D, or HOXB13 polypeptides to the exclusion of other polypeptides to produce a detectable signal. Recombinant, synthetic, and hybrid antibodies with the same ability may also be used in the practice of the invention. Antibodies may be readily generated by immunization with a IL17BR, CACNA1D, or HOXB13 polypeptide, and polyclonal sera may also be used in the practice of the invention.

20 Antibody based detection methods are well known in the art and include sandwich and ELISA assays as well as Western blot and flow cytometry based assays as non-limiting examples. Samples for analysis in such methods include any that contain IL17BR, CACNA1D, or HOXB13 polypeptides. Non-limiting examples include those containing breast cells and cell contents as well as bodily fluids (including blood, serum, saliva, lymphatic fluid, as well as mucosal and other
25 cellular secretions as non-limiting examples) that contain the polypeptides.

The above assay embodiments may be used in a number of different ways to identify or detect the response to treatment with TAM or another “antiestrogen” agent against breast cancer based on gene expression in a breast cancer cell sample from a patient. In some cases, this would reflect a secondary screen for the patient, who may have already undergone mammography or

physical exam as a primary screen. If positive from the primary screen, the subsequent needle biopsy, ductal lavage, fine needle aspiration, or other analogous minimally invasive method may provide the sample for use in the assay embodiments before, simultaneous with, or after assaying for ER status. The present invention is particularly useful in combination with non-invasive
5 protocols, such as ductal lavage or fine needle aspiration, to prepare a breast cell sample.

The present invention provides a more objective set of criteria, in the form of gene expression profiles of a discrete set of genes, to discriminate (or delineate) between breast cancer outcomes. In particularly preferred embodiments of the invention, the assays are used to discriminate between good and poor outcomes after treatment with tamoxifen or another
10 “antiestrogen” agent against breast cancer. Comparisons that discriminate between outcomes after about 10, about 20, about 30, about 40, about 50, about 60, about 70, about 80, about 90, about 100, or about 150 months may be performed.

While good and poor survival outcomes may be defined relatively in comparison to each other, a “good” outcome may be viewed as a better than 50% survival rate after about 60 months
15 post surgical intervention to remove breast cancer tumor(s). A “good” outcome may also be a better than about 60%, about 70%, about 80% or about 90% survival rate after about 60 months post surgical intervention. A “poor” outcome may be viewed as a 50% or less survival rate after about 60 months post surgical intervention to remove breast cancer tumor(s). A “poor” outcome may also be about a 70% or less survival rate after about 40 months, or about a 80% or less survival rate after
20 about 20 months, post surgical intervention.

In another embodiment of the invention based on the expression of a few genes, the isolation and analysis of a breast cancer cell sample may be performed as follows:

- (1) Ductal lavage or other non-invasive procedure is performed on a patient to obtain a sample.
- (2) Sample is prepared and coated onto a microscope slide. Note that ductal lavage results in
25 clusters of cells that are cytologically examined as stated above.
- (3) Pathologist or image analysis software scans the sample for the presence of atypical cells.
- (4) If atypical cells are observed, those cells are harvested (e.g. by microdissection such as LCM).
- (5) RNA is extracted from the harvested cells.

(6) RNA is assayed, directly or after conversion to cDNA or amplification therefrom, for the expression of IL17BR, CACNA1D, and/or HOXB13 sequences.

5 With use of the present invention, skilled physicians may prescribe or withhold treatment with TAM or another “antiestrogen” agent against breast cancer based on prognosis determined via practice of the instant invention.

The above discussion is also applicable where a palpable lesion is detected followed by fine needle aspiration or needle biopsy of cells from the breast. The cells are plated and reviewed by a pathologist or automated imaging system which selects cells for analysis as described above.

10 The present invention may also be used, however, with solid tissue biopsies, including those stored as an FFPE specimen. For example, a solid biopsy may be collected and prepared for visualization followed by determination of expression of one or more genes identified herein to determine the breast cancer outcome. As another non-limiting example, a solid biopsy may be collected and prepared for visualization followed by determination of HOXB13, IL17BR and/or
15 CACNA1D expression. One preferred means is by use of *in situ* hybridization with polynucleotide or protein identifying probe(s) for assaying expression of said gene(s).

In an alternative method, the solid tissue biopsy may be used to extract molecules followed by analysis for expression of one or more gene(s). This provides the possibility of leaving out the need for visualization and collection of only cancer cells or cells suspected of being cancerous.
20 This method may of course be modified such that only cells that have been positively selected are collected and used to extract molecules for analysis. This would require visualization and selection as a prerequisite to gene expression analysis. In the case of an FFPE sample, cells may be obtained followed by RNA extraction, amplification and detection as described herein.

25 In a further alternative to all of the above, the sequence(s) identified herein may be used as part of a simple PCR or array based assay simply to determine the response to treatment with TAM or another “antiestrogen” agent against breast cancer by use of a sample from a non-invasive or minimally invasive sampling procedure. The detection of sequence expression from samples may be by use of a single microarray able to assay expression of the disclosed sequences as well as other

sequences, including sequences known not to vary in expression levels between normal and non-normal breast cells, for convenience and improved accuracy.

Other uses of the present invention include providing the ability to identify breast cancer cell samples as having different responses to treatment with TAM or another “antiestrogen” agent against breast cancer for further research or study. This provides an advance based on objective genetic/molecular criteria.

The genes identified herein also may be used to generate a model capable of predicting the breast cancer survival and recurrence outcomes of an ER+ breast cell sample based on the expression of the identified genes in the sample. Such a model may be generated by any of the algorithms described herein or otherwise known in the art as well as those recognized as equivalent in the art using gene(s) (and subsets thereof) disclosed herein for the identification of breast cancer outcomes. The model provides a means for comparing expression profiles of gene(s) of the subset from the sample against the profiles of reference data used to build the model. The model can compare the sample profile against each of the reference profiles or against a model defining delineations made based upon the reference profiles. Additionally, relative values from the sample profile may be used in comparison with the model or reference profiles.

In a preferred embodiment of the invention, breast cell samples identified as normal and cancerous from the same subject may be analyzed, optionally by use of a single microarray, for their expression profiles of the genes used to generate the model. This provides an advantageous means of identifying survival and recurrence outcomes based on relative differences from the expression profile of the normal sample. These differences can then be used in comparison to differences between normal and individual cancerous reference data which was also used to generate the model.

Articles of Manufacture

The materials and methods of the present invention are ideally suited for preparation of kits produced in accordance with well known procedures. The invention thus provides kits comprising agents (like the polynucleotides and/or antibodies described herein as non-limiting examples) for the detection of expression of the disclosed sequences. Such kits, optionally comprising the agent

with an identifying description or label or instructions relating to their use in the methods of the present invention, are provided. Such a kit may comprise containers, each with one or more of the various reagents (typically in concentrated form) utilized in the methods, including, for example, pre-fabricated microarrays, buffers, the appropriate nucleotide triphosphates (e.g., dATP, dCTP, dGTP and dTTP; or rATP, rCTP, rGTP and UTP), reverse transcriptase, DNA polymerase, RNA polymerase, and one or more primer complexes of the present invention (e.g., appropriate length poly(T) or random primers linked to a promoter reactive with the RNA polymerase). A set of instructions will also typically be included.

The methods provided by the present invention may also be automated in whole or in part. All aspects of the present invention may also be practiced such that they consist essentially of a subset of the disclosed genes to the exclusion of material irrelevant to the identification of breast cancer survival outcomes via a cell containing sample.

Having now generally described the invention, the same will be more readily understood through reference to the following examples which are provided by way of illustration, and are not intended to be limiting of the present invention, unless specified.

Examples

Example 1

General methods

Patient and tumor selection criteria and study design

Patient inclusion criteria for this study were: Women diagnosed at the Massachusetts General Hospital (MGH) between 1987 and 2000 with ER positive breast cancer, treatment with standard breast surgery (modified radical mastectomy or lumpectomy) and radiation followed by five years of systemic adjuvant tamoxifen; no patient received chemotherapy prior to recurrence. Clinical and follow-up data were derived from the MGH tumor registry. There were no missing registry data and all available medical records were reviewed as a second tier of data confirmation.

All tumor specimens collected at the time of initial diagnosis were obtained from frozen and formalin fixed paraffin-embedded (FFPE) tissue repositories at the Massachusetts General Hospital.

Tumor samples with greater than 20% tumor cells were selected with a median of greater than 75% for all samples. Each sample was evaluated for the following features: tumor type (ductal vs. lobular), tumor size, and Nottingham combined histological grade. Estrogen and progesterone receptor expression were determined by biochemical hormone binding analysis and/or by immunohistochemical staining as described (Long, A.A. et al. "High-specificity in-situ hybridization. Methods and application." Diagn Mol Pathol 1, 45-57 (1992)); receptor positivity was defined as greater than 3 fmol/mg tumor tissue (Long et al.) and greater than 1% nuclear staining for the biochemical and immunohistochemical assays, respectively.

Study design is as follows: A training set of 60 frozen breast cancer specimens was selected to identify gene expression signatures predictive of outcome or response, in the setting of adjuvant tamoxifen therapy. Tumors from responders were matched to the non-responders with respect to TNM staging and tumor grade. Differential gene expression identified in the training set was validated in an independent group of 20 invasive breast tumors with formalin fixed paraffin-embedded (FFPE) tissue samples.

LCM, RNA isolation and amplification

With each frozen tumor sample within the 60-case cohort, RNA was isolated from both a whole tissue section of 8µm in thickness and a highly enriched population of 4,000-5,000 malignant epithelial cells acquired by laser capture microdissection using a PixCell Iie LCM system (Arcturus, Mountain View, CA). From each tumor sample within the 20-case test set, RNA was isolated from four 8µm-thick FFPE tissue sections. Isolated RNA was subjected to one round of T7 polymerase *in vitro* transcription using the RiboAmpTM kit (frozen samples) or another system for FFPE samples according to manufacturer's instructions (Arcturus Bioscience, Inc., Mountain View, CA for RiboAmpTM). Labeled cRNA was generated by a second round of T7-based RNA *in vitro* transcription in the presence of 5-[3-Aminoallyl]uridine 5'-triphosphate (Sigma-Aldrich, St. Louis, MO). Universal Human Reference RNA (Stratagene, San Diego, CA) was amplified in the same manner. The purified aRNA was later conjugated to Cy5 (experimental samples) or Cy3 (reference sample) dye (Amersham Biosciences).

Microarray analysis

A custom designed 22,000-gene oligonucleotide (60mer) microarray was fabricated using ink-jet in-situ synthesis technology (Agilent Technologies, Palo Alto, CA). Cy5-labeled sample RNA and Cy3-labeled reference RNA were co-hybridized at 65°C, 1X hybridization buffer (Agilent Technologies). Slides were washed at 37°C with 0.1X SSC/0.005% Triton X-102. Image analysis was performed using Agilent's image analysis software. Raw Cy5/Cy3 ratios were normalized using intensity-dependent non-linear regression.

A data matrix consisting of normalized Cy5/Cy3 ratios from all samples were median centered for each gene. The variance of each gene over all samples was calculated and the top 25% high variance genes (5,475) selected for further analysis. Identification and permutation testing for significance of differential gene expression were performed using BRB ArrayTools, developed by Dr. Richard Simon and Amy Peng (see [http site at linus.nci.nih.gov/BRB-ArrayTools.html](http://linus.nci.nih.gov/BRB-ArrayTools.html)). Hierarchical cluster analysis was performed with GeneMaths software (Applied-Maths, Belgium) using cosine correlation and complete linkage. All other statistical procedures (two-sample t-test, receiver operating characteristic analysis, multivariate logistic regression and survival analysis) were performed in the open source R statistical environment (see [http site at www.r-project.org](http://www.r-project.org)). Statistical test of significance of ROC curves was by the method of DeLong ("Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach." Biometrics 44, 837-45 (1988)). Disease free survival was calculated from the date of diagnosis. Events were scored as the first distant metastasis, and patients remaining disease-free at the last follow-up were censored. Survival curves were calculated by the Kaplan-Meier estimates and compared by log-rank tests.

Real-Time Quantitative PCR analysis

Real-time PCR was performed on 59 of the 60-case training samples (one case was excluded due to insufficient materials) and the 20-case validation samples. Briefly, 2 µg of amplified RNA was converted into double stranded cDNA. For each case 12ng of cDNA in triplicates was used for real-time PCR with an ABI 7900HT (Applied Biosystems) as described (Gelmini, S. et al. "Quantitative polymerase chain reaction-based homogeneous assay with fluorogenic probes to

measure c-erbB-2 oncogene amplification.” Clin Chem 43, 752-8 (1997)). The sequences of the PCR primer pairs and fluorogenic MGB probe (5' to 3'), respectively, that were used for each gene are as follows:

HoxB13

TTCATCCTGACAGTGGCAATAATC,
CTAGATAGAAAATATGAGGCTAACGATCAT,
VIC- CGATAACCAGTACTAGCTG;

IL17BR

GCATTAACCTAACGATTGGAACTACATT,
GGAAGATGCTTTATTGTTGCATTATC,
VIC-ACAACTTCAAAGCTGTTTTA.

Relative expression levels of HOXB13 in normal, DCIS and IDC samples were calculated as follows. First, all CT values are adjusted by subtracting the highest CT (40) among all samples, then relative expression = $1 / 2^{CT}$.

In Situ Hybridization

Dig-labeled RNA probes were prepared using DIG RNA labeling kit (SP6/T7) from Roche Applied Science, following the protocol provided with the kit. *In situ* hybridization was performed on frozen tissue sections as described (Long et al.).

Table 1. Patients and tumor characteristics of training set.

Sample ID	Tumor type	Size	Grade	Nodes	ER	PR	Age	DFS	Status
1389	D	1.7	2	0/1	Pos	Pos	80	94	0
648	D	1.1	2	0/15	Pos	ND	62	160	0
289	D	3	2	0/15	Pos	ND	75	63	1
749	D	1.8	2	2/9	Pos	Pos	61	137	0
420	D/L	2	3	ND	Pos	Pos	72	58	1
633	D	2.7	3	0/11	Pos	ND	61	20	1
662	D	1	3	6/11	Pos	Pos	79	27	1
849	D	2	1	0/26	Pos	Neg	75	23	1
356	D	1	2	2/20	Pos	ND	58	24	1
1304	D	2	3	0/14	Pos	Pos	57	20	1
1419	D	2.5	2	1/8	Pos	Pos	59	86	0

1093	D	1	3	1/14	Pos	Pos	66	85	0
1047	D/L	2.6	2	0/18	Pos	Neg	70	128	0
1037	D/L	1.5	2	0/4	Pos	Pos	85	83	0
319	D	4	2	1/13	Pos	ND	67	44	1
25	D	3.5	2	0/9	Neg	Pos	62	75	1
180	D	1.6	2	2/19	Pos	Pos	69	169	0
687	D	3.5	3	3/16	Pos	ND	73	142	0
856	D	1.6	2	0/16	Pos	Pos	73	88	0
1045	D	2.5	3	1/12	Pos	Neg	73	121	0
1205	D	2.7	2	1/19	Pos	Pos	71	88	0
1437	D	1.7	2	2/22	Pos	Pos	67	89	0
1507	D	3.7	3	0/40	Pos	Pos	70	70	0
469	D	1	1	0/19	Pos	ND	66	161	0
829	D	1.2	2	0/9	Pos	ND	69	136	0
868	D	3	3	0/13	Pos	Pos	65	130	0
1206	D	4.1	3	0/15	Pos	Neg	84	56	1
843	D	3.4	2	11/20	Pos	Neg	76	122	1
342	D	3	2	9/21	Pos	ND	62	102	1
1218	D	4.5	1	3/16	Pos	Pos	62	10	1
547	D/L	1.5	2	ND	Pos	ND	74	129	1
1125	D	2.6	2	0/18	Pos	Pos	54	123	0
1368	D	2.6	2	ND	Pos	Pos	82	63	0
605	D	2.2	2	6/18	Pos	ND	70	110	0
59	L	3	2	33/38	Pos	ND	70	21	1
68	D	3	2	0/17	Pos	ND	53	38	1
317	D	1.2	3	1/10	Pos	Pos	71	5	1
374	D	1	3	0/15	Pos	Neg	57	47	1
823	D	2	2	0/6	Pos	Pos	51	69	1
280	D	2.2	3	0/12	Pos	ND	66	44	1
651	D	4.7	3	10/13	Pos	ND	48	137	1
763	D	1.8	2	0/14	Pos	Pos	63	118	0
1085	D	4.7	2	0/8	Pos	Pos	48	101	1
1363	D	2.1	2	0/15	Pos	Pos	56	114	0
295	D	3.5	2	3/21	Pos	Pos	52	118	1
871	D	4	3	0/16	Pos	Neg	61	6	1
1343	D	2.5	3	ND	Pos	Pos	79	21	1
140	L	>2.0	2	18/28	Pos	ND	63	43	1
260	D/L	0.9	2	1/13	Pos	ND	73	42	1
297	D	0.8	2	1/16	Pos	Pos	66	169	0
1260	D	3.5	2	0/14	Pos	Pos	58	79	0
1405	D	1	3	ND	Pos	Pos	81	95	0
518	L	5.5	2	3/20	Pos	ND	68	156	0
607	D	1.2	2	5/14	Pos	Pos	76	114	0
638	D	2	2	1/24	Pos	Pos	67	148	0
655	D	2	3	ND	Pos	Pos	73	143	0

772	D	2.5	2	0/18	Pos	Pos	68	69	1
878	D/L	1.6	2	0/9	Pos	Neg	76	138	0
1279	D	2	2	0/12	Pos	Pos	68	102	0
1370	D	2	2	ND	Pos	Pos	73	61	0

Abbreviations: D, ductal; L, lobular; D/L, ductal and lobular features; pos, positive; neg, negative; ND, not determined; ER, estrogen receptor; PR, progesterone receptor; DFS, disease-free survival (number of months); status=1, recurred; status=0, disease-free.

5

Example 2

Identification of differentially expressed genes

Gene expression profiling was performed using a 22,000-gene oligonucleotide microarray as described above. In the initial analysis, isolated RNA from frozen tumor-tissue sections taken from the archived primary biopsies were used. The resulting expression dataset was first filtered based on overall variance of each gene with the top 5,475 high-variance genes (75th percentile) selected for further analysis. Using this reduced dataset, t-test was performed on each gene comparing the tamoxifen responders and non-responders, leading to identification of 19 differentially expressed genes at the P value cutoff of 0.001 (Table 2). The probability of selecting this many or more differentially expressed genes by chance was estimated to be 0.04 by randomly permuting the patient class with respect to treatment outcome and repeating the t-test procedure 2,000 times. This analysis thus demonstrated the existence of statistically significant differences in gene expression between the primary breast cancers of tamoxifen responders and non-responders.

Table 2. 19-gene signature identified by t-test in the Sections dataset

	Parametric p-value	Mean in responders	Mean in non-responders	Fold difference of means	GB acc	Description
1	1.96E-05	0.759	1.317	0.576	<u>AW006861</u>	SCYA4 small inducible cytokine A4
2	2.43E-05	1.31	0.704	1.861	<u>AI240933</u>	ESTs
3	8.08E-05	0.768	1.424	0.539	<u>X59770</u>	IL1R2 interleukin 1 receptor, type II
4	9.57E-05	0.883	1.425	0.62	<u>AB000520</u>	APS adaptor protein with pleckstrin homology and src homology 2 domains
5	9.91E-05	1.704	0.659	2.586	<u>AF208111</u>	IL17BR interleukin 17B receptor
6	0.0001833	0.831	1.33	0.625	<u>AI820604</u>	ESTs
7	0.0001935	0.853	1.459	0.585	<u>AI087057</u>	DOK2 docking protein 2, 56kD
8	0.0001959	1.29	0.641	2.012	<u>AJ272267</u>	CHDH choline dehydrogenase

9	0.0002218	1.801	0.943	1.91	N30081	ESTs, Weakly similar to I38022 hypothetical protein [H.sapiens]
10	0.0004234	1.055	2.443	0.432	AI700363	ESTs
11	0.0004357	0.451	1.57	0.287	AL117406	ABCC11 ATP-binding cassette, sub-family C (CFTR/MRP), member 11
12	0.0004372	1.12	3.702	0.303	BC007092	HOXB13 homeo box B13
13	0.0005436	0.754	1.613	0.467	M92432	GUCY2D guanylate cyclase 2D, membrane (retina-specific)
14	0.0005859	1.315	0.578	2.275	AL050227	Homo sapiens mRNA; cDNA DKFZp586M0723 (from clone DKFZp586M0723)
15	0.000635	1.382	0.576	2.399	AW613732	Homo sapiens cDNA FLJ31137 fis, clone IMR322001049
16	0.0008714	0.794	1.252	0.634	BC007783	SCYA3 small inducible cytokine A3
17	0.0008912	2.572	1.033	2.49	X81896	C11orf25 chromosome 11 open reading frame 25
18	0.0009108	0.939	1.913	0.491	BC004960	MGC10955 hypothetical protein MGC10955
19	0.0009924	1.145	0.719	1.592	AK027250	Homo sapiens cDNA: FLJ23597 fis, clone LNG15281

To refine our analysis to the tumor cells and circumvent potential variability attributable to stromal cell contamination, the same cohort was reanalyzed following laser-capture microdissection (LCM) of tumor cells within each tissue section. Using variance based gene filtering and t-test screening identical to that utilized for the whole tissue section dataset, 9 differentially expressed gene sequences were identified with $P < 0.001$ (Table 3).

Table 3. 9-gene signature identified by t-test in the LCM dataset

	Parametric p-value	Mean in responders	Mean in non-responders	Fold difference of means	GB acc	Description
1	2.67E-05	1.101	4.891	0.225	BC007092	HOXB13 homeo box B13
2	0.0003393	1.045	2.607	0.401	AI700363	ESTs
3	0.0003736	0.64	1.414	0.453	NM_014298	QPRT quinolate phosphoribosyltransferase (nicotinate-nucleotide pyrophosphorylase (carboxylating))
4	0.0003777	1.642	0.694	2.366	AF208111	IL17BR interleukin 17B receptor
5	0.0003895	0.631	1.651	0.382	AF033199	ZNF204 zinc finger protein 204
6	0.0004524	1.97	0.576	3.42	AI688494	FLJ13189 hypothetical protein FLJ13189
7	0.0005329	1.178	0.694	1.697	AI240933	ESTs
8	0.0007403	0.99	1.671	0.592	AL157459	Homo sapiens mRNA; cDNA DKFZp434B0425 (from clone DKFZp434B0425)

9	0.0007739	0.723	1.228	0.589	BC002480	FLJ13352 hypothetical protein FLJ13352
---	-----------	-------	-------	-------	----------	---

Only 3 genes were identified as differentially expressed in both the LCM and whole tissue section analyses: the homeobox gene HOXB13 (identified twice as AI700363 and BC007092), the interleukin 17B receptor IL17BR (AF208111), and the voltage-gated calcium channel CACNA1D (AI240933). HOXB13 was differentially overexpressed in tamoxifen nonresponsive cases, whereas IL17BR and CACNA1D were overexpressed in tamoxifen responsive cases. Based on their identification as tumor-derived markers significantly associated with clinical outcome in two independent analyses, the utility of each of these genes was evaluated by itself and in combination with the others.

To define the sensitivity and specificity of HOXB13, IL17BR and CACNA1D expression as markers of clinical outcome, Receiver Operating Characteristic (ROC) analysis (Pepe, M.S. "An interpretation for the ROC curve and inference using GLM procedures." *Biometrics* 56, 352-9 (2000)) was used. For data derived from whole tissue sections, the Area Under the Curve (AUC) values for IL17BR, HOXB13 and CACNA1D were 0.79, 0.67 and 0.81 for IL17BR, HOXB13 and CACNA1D, respectively (see Table 4 and Fig. 1, upper portion). ROC analysis of the data generated from the microdissected tumor cells yielded AUC values of 0.76, 0.8, and 0.76 for these genes (see Table 4 and Fig. 1, lower portion).

Table 4. ROC analysis of using IL17BR, CACNA1D and HOXB13 expression to predict tamoxifen response

	Tissue Sections		LCM	
	AUC	P value	AUC	P value
IL17BR	0.79	1.58E-06	0.76	2.73E-05
CACNA1D	0.81	3.02E-08	0.76	1.59E-05
HOXB13	0.67	0.012	0.79	9.94E-07
ESR1	0.55	0.277	0.63	0.038
PGR	0.63	0.036	0.63	0.033
ERBB2	0.69	0.004	0.64	0.027
EGFR	0.56	0.200	0.61	0.068

AUC, area under the curve; *P* values are AUC > 0.5.

A statistical test of significance indicated that these AUC values are all significantly greater than 0.5, the expected value from the null model that predicts clinical outcome randomly. Therefore, these three genes have potential utility for predicting clinical outcome of adjuvant tamoxifen therapy. As comparison, markers that are currently useful in evaluating the likelihood of response to tamoxifen were analyzed in comparison. The levels of ER (gene symbol ESR1) and progesterone receptor (PR, gene symbol PGR) are known to be positively correlated with tamoxifen response (see Fernandez, M.D., et al. "Quantitative oestrogen and progesterone receptor values in primary breast cancer and predictability of response to endocrine therapy." Clin Oncol 9, 245-50 (1983); Ferno, M. et al. "Results of two or five years of adjuvant tamoxifen correlated to steroid receptor and S-phase levels." South Sweden Breast Cancer Group, and South-East Sweden Breast Cancer Group. Breast Cancer Res Treat 59, 69-76 (2000); Nardelli, G.B., et al. "Estrogen and progesterone receptors status in the prediction of response of breast cancer to endocrine therapy (preliminary report)." Eur J Gynaecol Oncol 7, 151-8 (1986); and Osborne, C.K., et al. "The value of estrogen and progesterone receptors in the treatment of breast cancer." U 46, 2884-8 (1980)).

In addition, growth factor signaling pathways (EGFR, ERBB2) are thought to negatively regulate estrogen-dependent signaling, and hence contribute to loss of responsiveness to tamoxifen (see Dowsett, M. "Overexpression of HER-2 as a resistance mechanism to hormonal therapy for breast cancer." Endocr Relat Cancer 8, 191-5 (2001)). ROC analysis of these genes confirmed their correlation with clinical outcome, but with AUC values ranging only from 0.55 to 0.69, reaching statistical significance for PGR and ERBB2 (see Table 4).

The LCM dataset is particularly relevant, since EGFR, ERBB2, ESR1 and PGR are currently measured at the tumor cell level using either immunohistochemistry or fluorescence *in situ* hybridization. As individual markers of clinical outcome, HOXB13, IL17BR and CAC1D all outperformed ESR1, PGR, EGFR and ERBB2 (see Table 4).

Example 3

Identification of the HOXB13:IL17BR Expression Ratio

HOXB13:IL17BR expression ratio was identified as a robust composite predictor of outcome as follows. Since HOXB13 and IL17BR have opposing patterns of expression, the expression ratio of HOXB13 over IL17BR was examined to determine whether it provides a better composite predictor of tamoxifen response. Indeed, both t-test and ROC analyses demonstrated that the two-gene ratio had a stronger correlation with treatment outcome than either gene alone, both in the whole tissue sections and LCM datasets (see Table 5). AUC values for HOXB13:IL17BR reached 0.81 for the tissue sections dataset and 0.84 for the LCM dataset. Pairing HOXB13 with CACNA1D or analysis of all three markers together did not provide additional predictive power.

Table 5. HOXB13:IL17BR ratio is a stronger predictor of treatment outcome

		t-test		ROC	
		t-statistic	P value	AUC	P value
Tissue Section	IL17BR	4.15	1.15E-04	0.79	1.58E-06
	HOXB13	-3.57	1.03E-03	0.67	0.01
	HOXB13:IL17BR	-4.91	1.48E-05	0.81	1.08E-07
LCM	IL17BR	3.70	5.44E-04	0.76	2.73E-05
	HOXB13	-4.39	8.00E-05	0.79	9.94E-07
	HOXB13:IL17BR	-5.42	2.47E-06	0.84	4.40E-11

AUC, area under the curve; P values are AUC > 0.5.

15

The HOXB13/IL17BR ratio was compared to well-established prognostic factors for breast cancer, such as patient age, tumor size, grade and lymph node status (see Fitzgibbons, P.L. et al. "Prognostic factors in breast cancer. College of American Pathologists Consensus Statement 1999." Arch Pathol Lab Med 124, 966-78 (2000)). Univariate logistic regression analysis indicated that only tumor size was marginally significant in this cohort (P=0.04); this was not surprising given that the responder group was closely matched to the non-responder group with respect to tumor size, tumor grade and lymph node status during patient selection. Among the known positive (ESR1 and PGR) and negative (ERBB2 and EGFR) predictors of tamoxifen response, ROC analysis of the tissue sections data indicated that only PGR and ERBB2 were significant (see Table 4). Therefore,

25

a comparison of logistic regression models containing the HOXB13:IL17BR ratio either by itself or in combination with tumor size, and expression levels of PGR and ERBB2, were made (see Table 6). The HOXB13:IL17BR ratio alone was a highly significant predictor ($P=0.0003$) and had an odds ratio of 10.2 (95%CI 2.9-35.6). In the multivariate model, HOXB13:IL17BR ratio is the only significant variable ($P=0.002$) with an odds ratio of 7.3 (95%CI 2.1-26). Thus, the expression ratio of HOXB13:IL17BR is a strong independent predictor of treatment outcome in the setting of adjuvant tamoxifen therapy.

Table 6. Logistic Regression Analysis
Univariate Model

Predictor	Odds Ratio	95% CI	P Value
<i>HOXB13:IL17BR</i>	10.17	2.9-35.6	0.0003

Multivariate Model

Predictors	Odds Ratio	95% CI	P Value
<i>Tumor size</i>	1.5	0.7-3.5	0.3289
<i>PGR</i>	0.8	0.3-1.8	0.5600
<i>ERBB2</i>	1.7	0.8-3.8	0.1620
<i>HOXB13:IL17BR</i>	7.3	2.1-26.3	0.0022

All predictors are continuous variables. Gene expression values were from microarray measurements. Odds ratio is the inter-quartile odds ratio, based on the difference of a predictor from its lower quartile (0.25) to its upper quartile (0.75); CI, confidence interval.

Example 4

Independent validation of HOXB13:IL17BR expression ratio

The reduction of a complex microarray signature to a two-gene expression ratio allows the use of simpler detection strategies, such as real-time quantitative PCR (RT-QPCR) analysis. The HOXB13:IL17BR expression ratio by RT-QPCR using frozen tissue sections that were available from 59 of the 60 training cases were analyzed (Fig 2, part a). RT-QPCR data were highly concordant with the microarray data of frozen tumor specimens (correlation coefficient $r=0.83$ for HOXB13, 0.93 for IL17BR). In addition, the PCR-derived HOXB13:IL17BR ratios, represented as Δ CTs, where CT is the PCR amplification cycles to reach a predetermined threshold amount (e.g.,

Fig. 2, parts a and b) and ΔCT is the CT difference between HOXB13 and IL17BR, were highly correlated with the microarray-derived data ($r=0.83$) and with treatment outcome (t test $P=0.0001$, Fig. 2, part c). Thus, conventional RT-QPCR analysis for the expression ratio of HOXB13 to IL17BR appears to be equivalent to microarray-based analysis of frozen tumor specimens.

5 To validate the predictive utility of HOXB13:IL17BR expression ratio in an independent patient cohort, 20 additional ER-positive early-stage primary breast tumors from women treated with adjuvant tamoxifen only at MGH between 1991 and 2000, and for which medical records and paraffin-embedded tissues were available, were identified. Of the 20 archival cases, 10 had
10 recurred with a median time to recurrence of 5 years, and 10 had remained disease-free with a median follow up of 9 years (see Table 7 for details).

Table 7. Patient and tumor characteristics of the validation set.

Sample	Tumor Type	Size	Grade	Nodes	ER	PR	Age	DFS	Status
Test 1	D	1.9	3	0/10	Pos	Pos	69	15	1
Test 2	D	1.7	3	0/19	Pos	Pos	61	117	1
Test 3	D	1.7	2	0/26	Pos	ND	65	18	1
Test 4	D	1.2	2	0/19	Pos	Pos	63	69	1
Test 5	D	1.7	2	2/2	Pos	Pos	60	52	1
Test 6	D	1.1	1	0/10	Pos	Pos	54	59	1
Test 7	D	>1.6	2	0/17	Pos	Neg	66	32	1
Test 8	L	2.6	1-2	0/14	Pos	Pos	58	67	1
Test 9	D	1.2	2	ND	Pos	Pos	93	58	1
Test 10	D	4	3	0/20	Pos	Pos	66	27	1
Test11	D	1.1	2	0/19	Pos	Pos	64	97	0
Test 12	D	2.7	2	0/10	Pos	Pos	66	120	0
Test 13	D	0.9	1	0/22	Pos	Pos	66	123	0
Test 14	D	2.1	2	0/16	Pos	Pos	57	83	0
Test 15	D	0.8	1-2	0/8	Pos	Pos	74	80	0
Test 16	D	1	2	0/13	Pos	Pos	74	93	0
Test 17	D	1.6	2	0/29	Pos	Pos	66	121	0
Test 18	L	1.5	1-2	0/8	Pos	Pos	65	25	0
Test 19	D	1.5	3	0/16	Pos	Pos	60	108	0
Test 20*	L	4	1-2	0/19	Pos	Pos	60	108	0

Abbreviations: Same as supplemental Table 1. * Patient received tamoxifen

for 2 years.

RNA was extracted from formalin-fixed paraffin-embedded (FFPE) whole tissue sections, linearly amplified, and used as template for RT-QPCR analysis. Consistent with the results of the training cohort, the HOXB13:IL17BR expression ratio in this independent patient cohort was highly correlated with clinical outcome (t test $P=0.035$) with higher HOXB13 expression (lower ΔCTs) correlating with poor outcome (Fig. 2, part d). To test the predictive accuracy of the HOXB13:IL17BR ratio, the RT-QPCR data from the frozen tissue sections ($n=59$) was used to build a logistic regression model. In this training set, the model predicted treatment outcome with an overall accuracy of 76% ($P=0.000065$, 95% confidence interval 63%-86%). The positive and negative predictive values were 78% and 75%, respectively. Applying this model to the 20 independent patients in the validation cohort, treatment outcome for 15 of the 20 patients was correctly predicted (overall accuracy 75%, $P=0.04$, 95% confidence interval 51%-91%), with positive and negative predictive values of 78% and 73%, respectively.

Kaplan-Meier analysis of the patient groups as predicted by the model resulted in significantly different disease-free survival curves in both the training set and the independent test set (Fig. 2, parts e and f).

Additional References

Ma, X.J. et al. Gene expression profiles of human breast cancer progression. Proc Natl Acad Sci U S A 100, 5974-9 (2003).

Nicholson, R.I. et al. Epidermal growth factor receptor expression in breast cancer: association with response to endocrine therapy. Breast Cancer Res Treat 29, 117-25 (1994).

All references cited herein, including patents, patent applications, and publications, are hereby incorporated by reference in their entireties, whether previously specifically incorporated or not.

5 Having now fully described this invention, it will be appreciated by those skilled in the art that the same can be performed within a wide range of equivalent parameters, concentrations, and conditions without departing from the spirit and scope of the invention and without undue experimentation.

10 While this invention has been described in connection with specific embodiments thereof, it will be understood that it is capable of further modifications. This application is intended to cover any variations, uses, or adaptations of the invention following, in general, the principles of the invention and including such departures from the present disclosure as come within known or customary practice within the art to which the invention pertains and as may be applied to the essential features hereinbefore set forth.

Appendix

Sequences identified as those of IL17BR cluster

5 AW675096

CCGGCGATGTCGCTCGTGCTGCTAAGCCTGGCCGCGCTGTGCAGGAGCGCCGTACCCCGA
GAGCCGACCGTTCAATGTGGCTCTGAACTGGGCCATCTCCAGAGTGGATGCTACAACAT
GATCTAATCCCGGGAGACTTGAGGGACCTCCGAGTAGAACCTGTTACAACCTAGTGTGCA
10 ACAGGGGACTATTCAATTTTGATGAATGTAAGCTGGGTACTCCGGGCAGATGCCAGCATC
CGCTTGTTGAAGGCCACCAAGATTTGTGTGACGGGCAAAAGCAACTTCCAGTCTACAGC
TGTGTGAGGTGCAATTACACAGAGGCCTTCCAGACTCAGACCAGACCCTCTGGTGGTAAA
TGGACATTTTCTACATCGGCTTCCCTGTAGAGCTGAACACAGTCTATTTTCATTGGGGCC
CATAATATTCCTAATGCAAATATGAATGAAGATGGCCCTTCCATGTCTGTGAATNTCACC
TCACCAGGCTGCCTAGACCACATAATGAAATATAAAAAAAGTGTGTCAAGGCCGGAAGC
15 CTGTGGGATCCGAACATCACT

AW673932

TTTTTTTTTTTTTTTTTTTTTAAAGTGGGTTCAGCTTGTTTATTCCCTACTTTTGTTATCT
TAAAAACAATGATTTTTTGCATGTAATAGAAGGTTTTTCACTTAAGATGCTATTGAGTGA
20 ATCAGTGAGGGGTTCTTAGAGTTAGTATTCATTAATTAAACATAGAATATTAGCTAAACA
GTTCTGGGTACACTGCAATGCATGGTCTATGGAAGACTAGATGTTTGGCTGAAGATGCTT
TATTGTTGCATTATCAAAATGGTTATAGTTTTTCAATTAAACTGTAATTGATTTCTATGT
ATAAAACAGCTTTGAAGTTGTAAATGTAGTTTCCAATCGTTAGTTAATGCTACATTAGTT
AGCAATATTTGAAAATTTTATTGGTATAAAATGTTTTAATTACTAAGGCTGTTTGTAGGC
25 TGCATAGTAAGCTTCAGGATCATCACACGTTTTTTCCCTGTAATTGG

BC000980

ggcccggcga tgtcgctcgt gctgctaagc ctggccgcgc tgtgcaggag cgccgtaccc
cgagagccga ccgttcaatg tggctctgaa actgggccat ctccagagtg gatgctacaa
30 catgatctaa tcccgggaga cttgaggagc ctccgagtag aacctgttac aactagtgtt
gcaacagggg actattcaat tttgatgaat gtaagctggg tactccgggc agatgccagc
atccgcttgt tgaaggccac caagatttgt gtgacgggca aaagcaactt ccagtcctac
agctgtgtga ggtgcaatta cacagaggcc ttccagactc agaccagacc ctctggtggt
aatggacat tttcctacat cggcttcctt gtagagctga acacagtcta tttcattggg
35 gcccataata ttcctaataa aaatatgaat gaagatggcc cttccatgtc tgtgaatttc
acctcaccag gctgcctaga ccacataatg aaatataaaa aaaagtgtgt caaggccgga
agcctgtggg atccgaacat cactgcttgt aagaagaatg aggagacagt agaagtgaac
ttcacaacca ctcccctggg aaacagatac atggctctta tccaacacag cactatcatc
gggttttctc aggtgtttga gccacaccag aagaaacaaa cgcgagcttc agtggtgatt
40 ccagtgactg gggatagtga aggtgctacg gtgcagctga ctccatattt tcctacttgt

ggcagcgact gcatccgaca taaaggaaca gttgtgctct gccacaaac aggcgtccct
ttccctctgg ataacaacaa aagcaagccg ggaggctggc tgcctctcct cctgctgtct
ctgctggtgg ccacatgggt gctggtggca gggatctatc taatgtggag gcacgaaagg
atcaagaaga cttccttttc taccaccaca ctactgcccc ccattaaggt tcttgtgggt
5 taccatctg aaatatgttt ccatcacaca atttggtact tcaactgaatt tcttcaaaac
cattgcagaa gtgaggtcat ccttgaaaag tggcagaaaa agaaaatagc agagatgggt
ccagtgcagt ggcttgccac tcaaaagaag gcagcagaca aagtcgtctt ccttctttcc
aatgacgtca acagtgtgtg cgatgggtacc tgtggcaaga gcgagggcag tcccagttag
aactctcaag acctcttccc ccttgccctt aaccttttct gcagtgatct aagaagccag
10 attcatctgc acaaatacgt ggtggtctac tttagagaga ttgatacaaa agacgattac
aatgctctca gtgtctgccc caagtaccac ctcatgaagg atgccactgc tttctgtgca
gaacttctcc atgtcaagca gcaggtgtca gcaggaaaaa gatcacaagc ctgccacgat
ggctgctgct ccttgtagcc caccatgag aagcaagaga ccttaaaggc ttcctatccc
accaattaca gggaaaaaac gtgtgatgat cctgaagctt actatgcagc ctacaaacag
15 ccttagtaat taaaacattt tataccaata aaattttcaa atattgctaa ctaatgtagc
attaactaac gattggaaac tacatttaca acttcaaagc tgttttatac atagaaatca
attacagttt taattgaaaa ctataaccat tttgataatg caacaataaa gcattctcag
ccaaacatct agtcttccat agaccatgca ttgcagtgtc cccagaactg tttagctaat
attctatgtt taattaatga atactaactc taagaacccc tcaactgattc actcaatagc
20 atcttaagtg aaaaaccttc tattacatgc aaaaaatcat tgtttttaag ataacaaaag
tagggaataa acaagctgaa cccactttta aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa
aa

BI602183

25 AGCGGAGCTGCGGGTGGCCTGGATCCCGCGCAGTGGCCCCGGCGATGTCGCTCGTGCTGCT
AAGCCTGGCCACGCTGTGCAGGAGCGCCGTACCCGAGAGCCGACCGTTCAATGTGGCTC
TGAAACTGTGGACATTTTCCTATATCGGCTTCCCTGTAGAGCTGAAAACAGTCTATTTCA
TTGGGGCCCATAATATTCCTAATGCAAATATGAATGAAGATGGCCCTTCCATGTCTGTGA
ATTTACCTCACCAGGCTGCCTAGACCACATAATGAAATATAAAAAAAGTGTGTCAAGGC
30 CGGAAGCCTGTGGGATCCGAACATCACTGCTTGTAAGAAGAATGAGGAGACAGTAGAAGT
GAACTTCACAACCACTCCCCTGGGAAACAGATACATGGCTCATCCAACACAGCACTATCA
TCGGGTTTTCTCAGGTGTTTGAGCCACACCAGAAGAAACAAACGCGAGCTTCAGTGGTGA
TTCCAGTGACTGGGGATAGTGAAGGTGCTACGGTGCAGCTGACTCCATATTTTCCTACTT
GTGGCAGCGACTGCATCCGACATAAAGGAACAGTTGTGCTCTGCCCACAAACAGGCGTCC
35 CTTTCCCCTCTGGATAACAACAAAAGCAAGCCGGGAGGCTGGCTGCCTCTCCTCCTGCTG
TCTCTGCTGGTTGGCCACATTGGGTGCTGGTGGCAGGGATCTATCTAATGTGGAGGCACG
AAAGGATCCAGAAGACTTCCTTTTCTACCACAACTACTGCCCCATTAAGGTCCTGTGG
TTACCCATCTTGAAATATGTTCCCTCACACAATTTGTTACTTCACTGAATTCTTCAAAC
TG

40

BI458542

AGCGGAGCGTGCGGGTGGCCTGGATCCCGCGCAGTGGCCCCGGCGATGTCGCTCGTGCTGC

TAAGCCTGGCCACGCTGTGCAGGAGCGCCGTACCCCGAGAGCCGACCGTTCAATGTGGCT
CTGAAACTGTGGACATTTTCCTATATCGGCTTCCCTGTAGAGCTGAAAACAGTCTATTTT
ATTGGGGCCCATAAATATTCCTAATGCAAAATATGAATGAAGATGGCCCTTCCATGTCTGTG
AATTTACCTCACCAGGCTGCCTAGACCACATAATGAAATATAAAAAAAGTGTGTCAAG
5 GCCGGAAGCCTGTGGGATCCGAACATCACTGCTTGTAAGAAGAATGAGGAGACAGTAGAA
GTGAACTTCACAACCACTCCCCTGGGAAACAGATACATGGCTCATCCAACACAGCACTAT
CATCGGGTTTTCTCAGGTGTTTGAGCCACACCAGAAGAAACAAACGCGAGCTTCAGTGGT
GATTCCAGTGA CTGGGGATAGTGAAGGTGCTACGGTGCAGCTGACTCCATATTTTCCTAC
TTGTGGCAGCGACTGCATCCGACATAAAGGAACAGTTGTGCTCTGCCCACAAACAGGCGT
10 CCCTTTCCCTCTGGATAACAACAAAAGCAAGCCGGGAGGCTGGCTGCCTCTCCTCCTGCT
GTCTCTGCTGGTGGNCACATTGGGTGCTGGTGGCAGGGATCTATCTAATGTGGAGGCACG
AAAGGATCAGAAGACTTCCTTTTCTACCACCACATACTGCCCCCATTAAGGTTCTTGTG
GTTTACCC

15 BI823321

GGCGATGTCGCTCGTGCTGCTAAGCCTGGCCGCGCTGTGCAGGAGCGCCGTACCCCGAGA
GCCGACCGTTCAATGTGGCTCTGAAACTGGGCCATCTCCAGAGTGGATGCTACAACATGA
TCTAATCCCGGGAGACTTGAGGGACCTCCGAGTAGAACCTGTTACAAC TAGTGTGCAAC
AGGGGACTATTCAATTTTGATGAATGTAAGCTGGGTACTCCGGGCAGATGCCAGCATCCG
20 CTTGTTGAAGGCCACCAAGATTTGTGTGACGGGCAAAAGCAACTTCCAGTCCTACAGCTG
TGTGAGGTGCAATTACACAGAGGCCTTCCAGACTCAGACCAGACCCTCTGGTGGTAAATG
GACATTTTCTATATCGGCTTCCCTGTAGAGCTGAACACAGTCTATTTTCATTGGGGCCCA
TAATATTCTAATGCAAAATATGAATGAAGATGGCCCTTCCATGTCTGTGAATTTACCTC
ACCAGGAAGCCTGTGGGATCCGAACATCACTGCTTGTAAGAAAGAATGAGGAGACAGTAG
25 AAGTGAACCTTCACAACCACTCCCCTGGGAAACAGATACATGGCTCTTATCCAACACAGCA
CTATCATCGGGTTTTCTCAGGTGTTTGAGCCACACCAGAAGAAACAAACGCGAGCTTCAGT
GGTGATTCCAGTGA CTGGGGATAGTGAAGGTGCTACGGTGCAGCTGACTCCATATTTTCC
TACTTGTGGCAGCGACTGCAATCCGACATAAAGGAACAGTTGTGCTCTGCCCACAAACAG
GCGTCCCTTTCCCTCTTGATAGCAACAGAAGCAAGCCGGGAGGCTGGTGCCTCTTCTTC
30 TGGTGTCTCTGCTGGTGGCACATTGAGTGCTGGTGGCAGGATCCATCTAATGTGGAGGCC
CCAAAGGACCAGGAAAGACTTCCTTTATTAGCACCAAGTATTGCCC

AA514396

TGGCTGAAGATGCTTTATTGTTGCATTATCAAAATGGTTATAGTTTTCAATTAAACTGT
35 AATTGATTTCTATGTATAAAACAGCTTTGAAGTTGTAAATGTAGTTTCCAATCGTTAGTT
AATGCTACATTAGTTAGCAATATTTGAAAATTTTATTGGTATAAAATGTTTTAATTACTA
AGGCTGTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACACGTTTTTCCCTGTAATT
GGTGGGATAGGAAGCCTTTAAGGTCTCTTGCTTCTCATGGGTGGGCTACAAGGAGCAGCA
GCCATCGTGGCAGGCTTGTGATCTTTTTCTGCTGACACCTGCTACTTGACATGGAGAAG
40 TTCTGCACAGAAAGCAGTGGCATCCTTCATGAGGTGGTACTTGGGGCAGACACTGAGAGC
ATTGTAATCGTCTTTTGTATCAATCTCTCTAAAGTAGACCACCACGTATTTGTGCAGATG
AATCTGGC

BF110326

TTTGTGGCTGAAGATGCTTTATTGTTGCATTATCAAAATGGTTATAGTTTTCAATTAA
AACTGTAATTGATTTCTATGTATAAAACAGCTTTGAAGTTGTAAATGTAGTTTCCAATCG
5 TTAGTTAATGCTACATTAGTTAGCAATATTTGAAAATTTTATTGGTATAAAATGTTTTAA
TTACTAAGGCTGTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACACGTTTTTCCCT
GTAATTGGTGGGATAGGAAGCCTTTAAGGTCTCTTGCTTTTCATGGGTGGGCTACAAGGA
GCAGCAGCCATCGTGGCAGGCTTGTGATCTTTTTCTGCTGACACCTGCTGCTTGACATG
GAGAAGTTCTGCACAGAAAGCAGTGGCATCCTTCATGAGGTGGTACTTGGGGCAGACACT
10 GAGAGCATTGTAATCGTCTTTTGTATCAATCTCTCTAAAGTAGACCACCACGTATTTGTG
CAGATGAATCTGGCTTCTTAGATCACTGC

BE466508

TGGCATGAGATGCTATATTGTTGCATTATCAAAATGGGTTTAGTCTTCAATTAACACTGT
15 AATTGATTTCTATGTATAAAACAGCTTTGAAGTTGTAAATGTGGTTTCCAATCGTCAGTT
AATGCTACATTAGTTAGCAATATTTGAAAATTTTATTGGTATAAAATGTTTTAATTACTA
AGGCTGTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACACGTTTTTCCCTGTAATT
GGTGGGATAGGAAGCCTTTAAGGTCTCTTGCTTCTCATGGGTGGGCTACAAGGAGCAGCA
GCCATCGTGGCAGGCTTGTGATCTTTTTCTGCTGACACCTGCTGCTTGACATGGAGAAG
20 TTCTGCACAGAAAGCAGTGGCATCCTTCATGAGGTGGTACTTGGGGCAGACACTGAGAGC
ATTGTAATCGTCTTTTGTATCAATCTCTCTAAAGTAGACCACCACGTATTTGTGCAGATG
AATCTGGCTTCTTAGATCACTG

BF740045

GTTTGGCTGAAGATGCTTTATTGTTGCATTATCAAAATGGTTATAGTTTTCAATTAAAC
TGTAATTGATTTCTATGTATAAAACACGCTTTGAAGTTGTAAATGTAGTTTCCAATCGTT
AGTTAATGCTACATTAGTTAGCAATATTTGAAAATTTTATTGGTATAAAATGTTTTAATT
ACTAAGGCTGTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACACGTTTTTCCCTGT
AATTGGTGGGATAGGAAGCCTTTAAGGTCTCTTGCTTCTCATGGGTGGGCTACAAGGAGC
30 AGCAGCCATCGTGGCAGGCTTGTGATCTTTTTCTGCTGACACCTGCTGCTTGACATGGA
GAAGTTCTGCACAGAAAGCAGTGGCATCCTTCATGAGGTGGTACTTGGGGCAGACACTGA
GAGCATTGTAATCGTCTTTTGTATCAATCTCTCTAAAGTA

AW299271

TGGCTGAAGATGCTTTATTGTTGCATTATCAAAATGGTTATAGTTTTCAATTAAACTGT
AATTGATTTCTATGTATAAAACACGCTTTGAAGTTGTAAATGTAGTTTCCAATCGTTAGTT
AATGCTACATTAGTTAGCAATATTTGAAAATTTTATTGGTATAAAATGTTTTAATTACTA
AGGCTGTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACACGTTTTTCCCTGTAATT
GGTGGGATAGGAAGCCTTTAAGGTCTCTTGCTTCTCATGGGTGGGCTACAAGGAGCAGCA
40 GCCATCGTGGCAGGCTTGTGATCTTTTTCTGCTGACACCTGCTGCTTGACATGGAGAAG

TTCTGCACAGAAAGCAGTGGCATCCTTCATGAGGTGGTACTTGGGGCAGACACTGAGAGC
ATTGTAATCGTCTTTTGTATCAATCTCTCTAAAGTAGACCACCACGTATTTGTGCAGATG
AATCTGGCTTCTTAGATCACTGCAGAAAAG

5 AA836217

TTTTTTTTTTTACAACCTTCAAAGCTGTTTTATACATAGAAATCAATTACAGTTTTAATTGA
AAACTATAACCATTTTGTATAATGCAACAATAAAGCATCTTCAGCCAAACATCTAGTCTTC
CATAGACCATGCATTGCAGTGTACCCAGAAGCTGTTTAGCTAATATTCTATGTTTAATTAA
TGAATACTAACTCTAAGAACCCCTCACTGATTCACTCAATAGCATCTTAAGTGAAAAACC
10 TTCTATTACATGCAAAAAATCATTGTTTTTAAGATAACAAAAGTAGGGAATAACAAGCT
GAACCCACTTTTACTGGACCAAATGATCTATTATATGTGTACCACTTGTATGATTTGGTA
TTTGCATAAGACCTTCCCTCTACAAACTAGATTCATATCTTGATTCTTGTACAGGTGCCT
TTTAACATGAACAACAAAATACCCACAAACTTGTCTACTTTTGCC

15 AI203628

TAGTAATTAACACATTTTATACCAATAAAAATTTTCAAATATTGCTAACTAATGTAGCATT
AACTAACGATTGGAAACTACATTTACAACCTCAAAGCTGTTTTATACATAGAAATCAATT
ACAGTTTTTAATTGAAAACCTATAACCATTTTGTATAATGCAACAATAAAGCATCTTCAGCCA
AACATCTAGTCTTCCATAGACCATGCATTGCAGTGTACCCAGAAGCTGTTTAGCTAATATT
20 CTATGTTTAATTAATGAATACTAACTCTAAGAACCCCTCACTGATTCACTCAATAGCATC
TTAAGTGAAAAACCTTCTATTACATGCAAAAAATCATTGTTTTTAAGATAACAAAAGTAG
GGAATAACAAGCTGAACCCACTTTTACTGGACCAAATGATCTATTATATGTGTAACCAC
TTGTATGATTTGGTATTTGCATAAGACCTTCCCTCTACAAACTAGATTCATATCTTGATT
CTTGTACAGGTGCCTTTTAACATGAA

25

AI627783

TTTTTTTTTTTTTTTTTAGCAATATTTGAAAATTTTATTGGTATAAAAATGTTTTAATTACT
AAGGCTGTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACACGTTTTTCCCTGTAAT
TGGTGGGATAGGAAGCCTTTAAGGTCTCTTGCTTCTCATGGGTGGGCTACAAGGAGCAGC
30 AGCCATCGTGGCAGGCTTGTGATCTTTTTCCTGCTGACACCTGCTACTTGACATGGAGAA
GTTCTGCACAGAAAGCAGTGGCATCCTTCATGAGGTGGTACTTGGGGCAGACACTGAGAG
CATTGTAATCGTCTTTTGTATCAATCTCTCTAAAGTAGACCACCACGTATTTGTGCAGAT
GAATCTGGCTTCTTAGATCACTGCAGAAAAGGTTAAAGGCAAGGGGGAAGAGGTCTTGAG
AGTTCTC

35

AI744263

TTAAAGTGGGTTTCAGCTTGTTTATTCCCTACTTTTGTTATCTTAAAAACAATGATTTTTT
GCATGTAATAGAAGGTTTTTCACTTAAGATGCTATTGAGTGAATCAGTGAGGGGTCTTA
GAGTTAGTATTCAATTAATTAAACATAGAATATTAGCTAAACAGTTCTGGGTACACTGCAA
40 TGCATGGTCTATGGAAGACTAGATGTTTGGCTGAAGATGCTTTATTGTTGCATTATCAAA

ATGGTTACAGTTTTCAATTAAAGCTGTAATTGATTTCTATGTATAAAACAGCTTTGAAGT
TGTAATGTAGTTTCCAATCGTTAGTTAATGCTACATTAGTTAGCAATATTTGAAAATTT
TATTGGTATAAAATGTTTTAATTACTAAGGCTGTTTGTAGGCTGCATAGTAAGCTTCAGG
ATCATCACACGTTNTTTCCCTGTAATTGGTGGGATAGGAAGCCTTTA

5

AI401622

AGTTAGCAATATTTGAAAATTTTATTGGTATAAAATGTTTTAATTACTAAGGCTGTTTGT
AGGCTGCATAGTAAGCTTCAGGATCATCACACGTTTTTTCCCTGTAATTGGTGGGATAGG
AAGCCTTTAAGGTCTCTTGCTTCTCATGGGTGGGCTACAAGGAGCAGCAGCCATCGTGGC
10 AGGCTTGTGATCTTTTTCCCTGCTGACACCTGCTACTTGACATGGAGAAGTTCTGCACAGA
AAGCAGTGGCATCCTTCATGAGGTGGTACTTGGGGCAGACACTGAGAGCATTGTAATCGT
CTTTTGTATCAATCTCTCTAAAGTAGACCACCACGTATTTGTGCAGATGAATCTGGCTTC
TTAGATCACTGCAGAAAAGGTTAAAGGCAAGGGGGAAGAGGTCTTGAGAGTTCTCACTGG

15

AI826949

TTGGCTGAAGATGCTTTTATTGTTGCATTATCAAAATGGTTATAGTTTTCAATTAAAACTG
TAATTGATTTCTATGTATAAAACAGCTTTGAAGTTGTAAATGTAGTTTCCAATCGTTAGT
TAATGCTACATTAGTTAGCAATATTTGAAAATTTTATTGGTATAAAATGTTTTAATTACT
AAGGCTGTTTGTAGGCTTGATAGTAAGCTTCAGGATCATCACACGTTTTTTCCCTGTAAT
20 TGGTGGGATAGGAAGCCTTTAAGGTCTCTTGCTTCTCATGGGTGGGCTACAAGGAGCAGC
AGCCATCGTGGCAGGCTTGTGATCTTTTTCCCTGCTGACACCTGCTGCTTGACATGGAGAA
GTTCTGCACAGAAAGCAGTGGCATCCTTCATGAGGTGGTACTTGGGGCAGACACTGAGAG
CATTGTAATCGTCT

25

BE047352

TTTTTTTTTTAGCAATATTTGAAAATTTTATTGGTATAAAATGTTTTAATTACTAAGGCT
GTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACACGTTTTTTCCCTGTAATTGGTGG
GATAGGAAGCCTTTAAGGTCTCTTGCTTCTCATGGGTGGGCTACAAGGAGCAGCAGCCAT
CGTGGCAGGCTTGTGATCTTTTTCCCTGCTGACACCTGCTACTTGACATGGAGAAGTTCTG
30 CACAGAAAGCAGTGGCATCCTTCATGAGGTGGTACTTGGGGCAGACACTGAGAGCATTGT
AATCGTCTTTTGTATCAATCTCTCTAAAGTAGACCACCACGTATTTGTGCAGATGAATCT
GGCTTCTTAGATCACTGCAGAAAAGGTTAAAGGCAAGGGGGAAGAGGTCTTGAGAG

AI911549

TTTGGCTGAAGATGCTTTTATTGTTGCATTATCAAAATGGTTACAGTTTTCAATTAAAGCT
GTAATTGATTTCTATGTATAAAACAGCTTTGAAGTTGTAAATGTAGTTTCCAATCGTTAG
TTAATGCTACATTAGTTAGCAATATTTGAAAATTTTATTGGTATAAAATGTTTTAATTAC
TAAGGCTGTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACACGTTTTTTCCCTGTAA
TTGGTGGGATAGGAAGCCTTTAAGGTCTCTTGCTTCTCATGGGTGGGCTACAAGGAGCAG
40 CAGCCATCGTGGCAGGCTTGTGATCTTTTTCCCTGCTGACACCTGCTGCTTGACATGGAGA

AGTTCTGCACAGAAAGCAGTGGCATCCTTCATGAGGTGGTACTTGGGGCAGACA

BF194822

5 TTCTCTGGCTGAAGATGCTTTATTGTTGCATTATCAAAATGGTTACAGTTTTCAATTAAA
GCTGTAATTGATTTCTATGTATAAAACAGCTTTGAAGTTGTAAATGTAGTTTCCAATCGT
TAGTTAATGCTACATTAGTTAGCAATATTTGAAAATTTTATTGGTATAAAATGTTTTAAT
TACTAAGGCTGTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACACGTTTTTCCCTG
TAATTGGTGGGATAGGAAGCCTTTAAGGTCTCTTGCTTCTCATGGGTGGGCTACAAGGAG
10 CAGCAGCCATCGTGGCAGGCTTGTGATCTTTTTCTGCTGACACCTGCTGCTTGACATGG
AGAAGTTCTGCACAGAAAGCAGTGGCATCCTTCATGAGGTGGTACTTGG

AI034244

15 TTTTTTTTTTTTTTTTACAACCTTGAAAGCTGTTTTATACATAGAAATCAATTACAGTTT
TAATTGAAAACCTATAACCATTTTGATAATGCAACAATAAAGCATCTTCAGCCAAACATCT
AGTCTTCCATAGACCATGCATTGCAGTGTACCCAGAAGTGTAGCTAATATTCTATGTT
TAATTAATGAATACTAAGTCTAAGAACCCCTCACTGATTCACTCAATAGCATCTTAAGTG
AAAAACCTTCTATTACATGCAAAAAATCATTGTTTTTAAGATAACAAAAGTAGGGAATAA
ACAAGCTGAACCCACTTTTACTGGACCAAATGATCTATTATATGTGTAACCACTTGTATG
20 ATTTGGATTGTCATAAGACCTTCCCTCTACAACTAGATTCATATCTTGATTCT

AI033911

25 TTTTTTTTTTTTTTTTACAACCTGCAAAGCTGTTTTATACATAGAAATCAATTACAGTTTT
AATTGAAAACCTATAACCATTTTGATAATGCAACAATAAAGCATCTTCAGCCAAACATCTA
GTCTTCCATAGACCATGCATTGCAGTGTACCCAGAAGTGTAGCTAATATTCTATGTTT
AATTAATGAATACTAAGTCTAAGAACCCCTCACTGATTCACTCAATAGCATCTTAAGTGA
AAAACCTTCTATTACATGCAAAAAATCATTGTTTTTAAGATAACAAAAGTAGGGAATAAA
CAAGCTGAACCCACTTTTACTGGACCAAATGATCTATTATATGTGTAACCACTTGTATGA
TTTGGTATTTGCATAAGACCTTCCCTCTACAACTAGATTCATATCTTGATTCT

BF064177

30 TTTTTTAGTTAGCAATATTTGAAAATTTTATTGGTATAAAATGTTTTAATTACTAAGGCT
GTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACACGTTTTTCCCTGTAATTGGTGG
GATAGGAAGCCTTTAAGGTCTCTTGCTTCTCATGGGTGGGCTACAAGGAGCAGCAGCCAT
CGTGGCAGGCTTGTGATCTTTTTCTGCTGACACCTGCTACTTGACATGGAGAAGTTCTG
35 CACAGAAAGCAGTGGCATCCTTCATGAGGTGGTACTTGGGGCAGACACTGAGAGCATTGT
AATCGTCTTTTGTATCAATCTCTCTAAAGTAGACCACCACGTATTTGTGCAGATGAATCT
GGCTTCTTAGATCACTGCAGAAAAGGTTAAAGGCAAGGGGAAGAGGTCTTGAGAGTTCT
CACTGGGACTGCCCTCGCTCTTGCCACAGGTACCATCGCACACACTGTTGACGTCATTGG
40 AAAG

AA847767

GGCTGAAGATGCTTTATTGTTGCATTATCAAAATGGTTATAGTTTTCAATTAAACTGTA
ATTGATTTCTATGTATAAAACAGCTTTGAAGTTGTAAATGTAGTTTCCAATCGTTAGTTA
ATGCTACATTAGTTAGCAATATTTGAAAATTTTATTGGTATAAAATGTTTTAATTACTAA
5 GGCTGTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACACGTTTTTTCCCTGTAATTG
GTGGGATAGGAAGCCTTTAAGGTCTCTTGCTTCTCATGGGTGGGCTACAAGGAGCAGCAG
CCATCGTGCGCAGGCTTGTGATCTTTTTCCCTGCTGACACCTGCTGCTTGACATGGAGAAGT
TCTGCACAGAAAGCAGTGGCATCCTTCATGAGGTGGTA

10 AI538624

TTGGCTGAAGATGCTTTATTGTTGCATTATCAAAATGGTTACAGTTTTCAATTAAAGCTG
TAATTGATTTCTATGTATAAAACAGCTTTGAAGTTGTAAATGTAGTTTCCAATCGTTAGT
TAATGCTACATTAGTTAGCAATATTTGAAAATTTTATTGGTATAAAATGTTTTAATTACT
AAGGCTGTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACACGTTTTTTCCCTGTAAT
15 TGGTGGGATAGGAAGCCTTTAAGGTCTCTTGCTTCTCATGGGTGGGCTACAAGGAGCAGC
AGCCATCGTGCGCAGGCTTGTGATCTTTTTCCCTGCTGACACCTGCTGCTTGACATGGAGAA
GTTCTGCACAGAAAGCAGTGGCATCCTTCATGAGGTGGTAC

AI913613

20 TTGGCTGAAGATGCTTTATTGTTGCATTATCAAAATGGTTATAGTTTTCAATTAAACTG
TAATTGATTTCTATGTATAAAACAGCTTTGAAGTTGTAAATGTAGTTTCCAATCGTTAGT
TAATGCTACATTAGTTAGCAATATTTGAAAATTTTATTGGTATAAAATGTTTTAATTACT
AAGGCTGTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACACGTTNNTTCCCTGTAAT
TGGTGGGATAGGAAGCCTTTAAGGTCTCTTGCTTCTCATGGGTGGGCTACAAGGAGCAGC
25 AGCCATCGTGCGCAGGCTTGTGATCTTTTTCCCTGCTGACACCTGCTGCTTGACATGGAGAA
GTTCTGCACAGAAAGCAGTGGCATCCTTCATG

AI942234

30 GTTTGGCTGAAGATGCTTTATTGTTGCATTATCAAAATGGTTATAGTTTTCAATTAAAC
TGTAATTGATTTCTATGTATAAAACAGCTTTGAAGTTGTAAATGTAGTTTCCAATCGTTA
GTTAATGCTACATTAGTTAGCAATATTTGAAAATTTTATTGGTATAAAATGTTTTAATTA
CTAAGGCTGTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACACGTTTTTTCCCTGTA
ATTGGTGGGATAGGAAGCCTTTAAGGTCTCTTGCTTCTCATGGGTGGGCTACAAGGAGCA
GCAGCCATCGTGCGCAGCTTGGTGATCTTTTTCCCTGCTGACACCTGCTGCTTGACATGAAG
35 AAGTTCTGCACAGAAAGCAGTGGCAT

AI580483

40 GTTTGGCTGAAGATGCTTTATTGTTGCATTATCAAAATGGTTATAGTTTTCAATTAAAC
TGTAATTGATTTCTATGTATAAAACAGCTTTGAAGTTGTAAATGTAGTTTCCAATCGTTA
GTTAATGCTACATTAGTTAGCAATATTTGAAAATTTTATTGGTATAAAATGTTTTAATTA

CTAAGGCTGTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACACGTTTTTCCCTGTA
ATTGGTGGGATAGGAAGCCTTTAAGGTCTCTTGCTTCTCATGGGTGGGCTACAAGGAGCA
GCAGCCATCGTGGCAGGCTTGGATCTTTTTCCTGCTGACACCTGCTGCTTGACATTGGAA
AGTTCTGCACAGAAAGCAGTGGCATC

5

AI831909

TTTTGGCTGATGATGCTTTATTGTTGCATTATCAAAATGGTTACAGTTTTCAATTAAAGC
TGTAATTGATTTCTATGTATAAAACAGCTTTGAAGTTGTAAATGTAGTTTCCAATCGTTA
GTTAATGCTACATTAGTTAGCAATATTTGAAAATTTTATTGGTATAAAATGTTTAAATTA
CTAAGGCTGTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACACGTTTTTCCCTGTA
ATTGGTGGGATAGGAAGCCTTTAAGGTCTCTTGCTTCTCATGGGTGGGCTACAAGGAGCA
GCAGCCATCGTGGCAGGCTTGTGATCTTTTTCCTGCTGACACCTGCTGCTTGACATGGAG
AAGTTCTGCACAGAAAGCAGTGGCAT

10

15

AI672344

GGCTGAAGATGCTTTATTGTTGCATTATCAAAATGGTTACAGTTTTCAATTAAAGCTGTA
ATTGATTTCTATGTATAAAACAGCTTTGAAGTTGTAAATGTAGTTTCCAATCGTTAGTTA
ATGCTACATTAGTTAGCAATATTTGAAAATTTTATTGGTATAAAATGTTTAACTACTAA
GGCTGTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACACGTTTTTCCCTGTAATTG
GTGGGATAGGAAGCCTTTAAGGTCTCTTGCTTCTCATGGGTGGGCTACAAGGAGCAGCAG
CCATCGTGGCAGGCTTGTGATCTTTTTCCTGCTGACACCTGCTGCTTGACATGGAGAAGT
TCTGCACAGAAAG

20

AW025192

GATTGGCTGTTTTATGCTTTATTGTTGCATTATCAAAATGGTTATAGTTTTCAATTAAAA
CTGTAATTGATTTCTATGTATAAAACAGCTTTGAAGTTGTAAATGTAGTTTCCAATCGTT
AGTTAATGCTACATTAGTTAGCAATATTTGAAAATTTTATTGGTATAAAATGTTTAAATT
ACTAAGGCTGTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACACGTTTTTCCCTGT
TATTGGTGGGATAGGAAGCCTTTAAGGTCTCTTGCTTCTCATGGGTGGGCTACAAGGAGC
AGCAGCCATCGTGGCAGGCTTGTGATCTTTTTCCTGCTGACACCTGCTGCTTGACATGGA
GAAGTTCTGCACAAAAAGCAGTGGCATCCTTCATGAGGTGGTA

25

30

AA677205

GCAATATTTTAAATTTTATTGGTATAAAATGTTTAACTAAGGCTGTTTGTAGGCT
GCATAGTAAGCTTCAGGATCATCACACGTTTTTCCCTGTAATTGGTGGCATAGGAAGCC
TTTAAGGTCTCTTGCTTCTCATGGTGTGGGCTACAAGGAGCAGCAGCCATCGTGGCAGGC
TTGTGATCTTTTTCCTGCTGACACCTGCTGCTTGACATGGAGAAGTTCTGCACAGAAAGC
AGTGGCATCCTTCATGAGGTGGTACTTGGGGCAGACACTGAGAGCATTGTAATCGTCTTT
TGTATCAATCTCTCTAAAGTAGACCACCGTATTTGTGCAGATGAATCTGGCTTCTTAG
ATCACTGCAGAAAAGGTTAAAGGCAAGGGGAAGAGGTCTTGAGAGTTCTCACTGGGACT

35

40

GCCCTCGCTCTTGCCACAGGTACCATCGCACACACTG

AA721647

TTTTTTTTTTTACAACCTTCAAAGCTGTTTTATACATAGAAATCAATTACAGTTTTAATTGA
5 AAACCTATAACCATTTTGATAATGCAACAATAAAGCATCTTCAGCCAAACATCTAGTCTTC
CATAGACCATGCATTGCAGTGTACCCAGAAGCTGTTTAGCTAATATTCTATGTTTAATTAA
TGAATACTAACTCTAAGAACCCCTCACTGATTCACTCAATAGCATCTTAAGTGAAAAACC
TTCTATTACATGCAAAAAATCATTGTTTTTAAGATAACAAAAGTAGGGAATAAACAAGCT
10 GAACCCACTTTTACTGGACCAAATGATCTATTATATGTGTAACCACTTGTATGATTTGGT
ATTTG

BF115018

GTTTCGCTGAAGATGCTTTATTGTTGCATTATCAAAATGGTTATAGTTTTCAATTAAAAC
TGTAATTGATTTCTATGTATAAAACAGCTTTGAAGTTGTAAATGTAGTTTCCAATCGTTA
15 GTTAATGCTACATTAGTTAGCAATATTTGAAAATTTTATTGGTATAAAATGTTTTAATTA
CTAAGGCTGTTTGTAGGCTGCATAGTAAGCTTAAGGCCCATCACACGTTTTTTCCCTGTA
ATTGGTGGGATAGGAAGCCTTTAAGGTCTCTTGCTTNTCATGGGTGGGCTACAAGGAGCA
GCAGCCATCGTGGCAGGCTTGNGATCTTTTTCTGCTGGCCCCTGCTGCTTGACAT

20 W61238

NAAAGCACTGGCTGAAGGAAGCCAAGAGGATCACTGCTGCTCCTTTTTTCTAGAGGAAAT
GTTTGTCTACGTGGTAAGATATGACCTAGCCCTTTTAGGTAAGCGAACTGGTATGTTAGT
AACGTGTACAAAGTTTtaggttcagacccccgggagtccttgggcacgtgggtctcgggtcac
TGGTTTTGACTTTAGGGCTTTGTTACAGATGTGTGACCAAGGGGAAAATGTGCATGACAA
25 CACTAGAGGTATGGGCGACACGANAAACGAACGGGAAGTTTTGGCTGAAGTAGGAGTCTTG
GTGAGATTTTGCTCTGATGCATGGTGTGAACTTTCTGAGCCTCTTGTTTTTCCTCAAGCT
GACTCCATATTTTCCTACTTGTGGCAGCGACTGCATCCGACATAAAGGAACAG

W61239

30 TAGCAATATTTGAAAATTTTATTGGTATAAAATGTTTTAATTACTAAGGCTGTTTGTAGG
CTGCATAGTAAGCTTCAGGATCATCACACGTTTTTTCCCTGTAATTGGTGGGATAGGAAG
CCTTTAAGGTCTCTTGCTTCTCATGGGTGGGCTACAAGGAGCAGCAGCCATCGTGGCAGG
CTTGTGATCTTTTTTCCTGCTGACACCTGCTGCTTGACATGGAGAAGTTCTGCACAGAAAG
CAGTGGCATCCTTCATGAGGTGGTACTTGGGGCAGACACTGAGAGCATTGTAATCGTCTT
35 TTGTATCAATCTCTCTAAAGTAGACCACCACGTATTTGTGCAGATGAATCTGGCTTCTTA
GATCACTGCAGAAAAGGTTAAAGGCAAGGGGGGA

AI032064

AGCAATATTTGAAAATTTTATTGGTATAAAATGTTTTAATTACTAAGGCTGTTTGTAGGC

TGCATAGTAAGCTTCAGGATCATCACACGTTTTTCCCTGTAATTGGTGGCATAGGAAGC
CTTTAAGGTCTCTTGCTTCTCATGGGTGGGCTACAAGGAGCAGCAGCCATCGTGGCAGGC
TTGTGATCTTTTTCCCTGCTGACACCTGCTGCTTGACATGGAGAAGTTCTGCACAGAAAGC
AGTGGCATCCTTCATGAGGTGGTACTTGGGGCAGACACTGAGAGCATTGTAATCGTCTTT
5 TGTATCAATCTCTCTAAAGTAGACCACCACGTATTTGTGCAGATGAATCTGGCTTCTTAG
ATCACTGCAGAAAAGGTTAAAGGCAAGGGGAAGAGGTCTTGAGAGTTCTCACTGGGACT
GCCCTCGCTCTTGCCAC

AW236941

10 TTTTTTTTTTTAGCAATATTTGAAAATTTTATTGGTATAAAATGTTTTAATTACTAAGGC
TGTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACACGTTTTTCCCTGTAATTGGTG
GGATAGGAAGCCTTTAAGGTCTCTTGCTTCTCATGGGTGGGCTACAAGGAGCAGCAGCCA
TCGTGGCAGGCTTGTGATCTTTTTCCCTGCTGACACCTGCTGCTTGACATGGAGAAGTTCT
GCACAAAAGCAGTGGCATCCTTCATGAGGTGGTACTTGGGGCAGACACTGAGAGCATTG
15 TAATCGTCTTTTGTATCAATC

AW236941

TTTTTTTTTTTTAGCAATATTTGAAAATTTTATTGGTATAAAATGTTTTAATTACTAAGGC
TGTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACACGTTTTTCCCTGTAATTGGTG
20 GGATAGGAAGCCTTTAAGGTCTCTTGCTTCTCATGGGTGGGCTACAAGGAGCAGCAGCCA
TCGTGGCAGGCTTGTGATCTTTTTCCCTGCTGACACCTGCTGCTTGACATGGAGAAGTTCT
GCACAAAAGCAGTGGCATCCTTCATGAGGTGGTACTTGGGGCAGACACTGAGAGCATTG
TAATCGTCTTTTGTATCAATC

BG057174

25 TTTTATACATAGAAATCAATTACAGCTTTAATTGAAAACCTATAACCATTTTGATAATGCA
ACAATAAAGCATCTTCAGCCAAACATCTAGTCTTCCATAGACCATGCATTGCAGTGTACC
CAGAACTGTTTAGCTAATATTCTATGTTTAATTAATGAATACTAACTCTAAGAACCCCTC
ACTGATTCACTCAATAGCATCTTAAGTGAAAAACCTTCTATTACATGCAAAAAATCATTG
30 TTTTAAAGATAACAAAAGTAGGGAATAACAAGCTGAACCCACTTTTACTGGACCAAATG
ATCTATTATATGTG

AW058532

GGCTGAAGATGCTTTATTGTTGCATTATCAAAATGGTTATAGTTTTCAATTAAAACTGTA
35 ATTGATTTCTATGTATAAAACAGCTTTGAAGTTGTAAATGTAGTTTCCAATCGTTAGTTA
ATGCTACATTAGTTAGCAATATTTGAAAATTTTATTGGTATAAAATGTTTTAATTACTAA
GGCTGTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACACGTTTTTCCCTGTATGG
GTGGGATAGGAAGCCTTTAAGGTCTCTTGCTTCTCATGGGTGGGCT

T98360

TNAGGAANGAGAAGAAGCGAGATNNANNTNNAGAAATANGTGGTGGCNTANTTTAGAGAG
ATTGATNCAAAAGCNGATTNCAATNNNCTCAGTGNCNTNCCCAAGTNCCNCCTCATGAAGG
ATNCACTNCTTTCTGTGCAGACTNNNCATGTCAAGCAGCAGGTGTCAGCAGGAAAAAGAN
CACAAGCTCCNCGATGGCTGCTGCTCCTTGTAGCCCNCCATGAGAAGCAAGAGNCTTAAA
5 GGCTTCCTATCCCACCAATTACAGGGAAAAACGTGTGATGACCTGAGCTTACTATGCAGC
CTACAANCAGCCTTAGTAATTAAACCNTTTATT

T98361

NANNATGAAGATGCTTTATTGTTGCATTATCAAAATGGTTACAGTTTTCAATTAAAGCTG
10 TAATTGATTTCTATGTATAAAACAGCTTTGAAGTTGTAAATGTAGTTTCCAATCGTTAGT
TAATGCTACATTAGTTAGCAATATTTGAAAATTTTATTGGNATAAAATGTTTTAATTACT
AAGGCTGTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACACGTTTTTNCCCTGTAAT
TGGGTGGGGATAGGGAAGCCCTTTAAGGGTCTCTTGCTTCTCATGGGGTGGGGCCTACNA
AGGGAGCAGCCAGCCCATCGTGGCCAGGGCCTTGTGGANCCTTTTTCCCTGCCTGGACAC
15 CCTGCCTGCCTTGGACCATGGGGAGGAAGGTTCTGGCACCAGGAAAGCCAGGTGGCCCAT
CCCTTCCATGAGGGTGGGGTACTTNGGGGGGCCAGGACCACTGAGGNGCCATTGGTAATC
CGTCCTTTTNGTATCCAATCCCCTCCTAAGGTAGGNCCCCC

AI470845

TTTTGTGGGTTTCAGCTTGTTTATTCCCTACTTTTGTTATCTTAAAAACAATGATTTTTTG
20 CATGTAATAGAAGGTTTTTCACTTAAGATGCTATTGAGTGAATCAGTGAGGGGTTCTTAG
AGTTAGTATTCATTAATTAACATAGAATATTAGCTAAACAGTTCTGGGTACACTGCAAT
GCATGGTCTATGGAAGACTAGATGTTTGGCTGAAGATGCTTTTATTGTTGCATTATCAAN
ATGGTTTATAGTTTTCAATTAAAACTGTAATTGATTT

25

AI497731

GGCTGAAGATGCTTTATTGTTGCATTATCAAAATGGTTATAGTTTTCAATTAAAACTGTA
ATTGATTTCTATGTATAAAACAGCTTTGAAGTTGTAAATGTAGTTTCCAATCGTTAGTTA
ATGCTACATTAGTTAGCAATATTTGAAAATTTTATTGGTATAAAATGTTTTAATTACTAA
30 GGCTGTTTGTAGGCTGCATAGTAAGCTTAANGATCATACNCACGTTTTTCCCTGAATTTG
GTGGGATAANGAAGCCTTTAAAGGT

T96629

TTGAAAATTTTATTGGNATAAAATGTTTTAATTACTAAGGCTGTTTGTAGGCTGCATAGT
35 AAGCTTCAGGANCATCACACGTTTTTTCCCTGTAATTGGTGGCATAGGAAGCCTTTAAGG
TCTCTTGCTTCTCATGGGTGGGCTACAAGGAGCAGCAGCCATCGTGGCAGGCTTGTGATC
TTTTTCCCTGCTGACACCTGCTGCTTGACATGGAGAAGTTCTGCACAGAAAGCAGTGGCAT
CCTTCATGAGGTGGTACTTGGGGCAGACACTGAGAGCATTGTAATCGTCTTTTGTATCAA
TCTCTCTAAAGTAGACCACCACCGTNTTTGTGCAGATGGANTCTGGCTTC

40

T96740

AGGCACTATCATCGGGTTTTCTCAGGTGTTTGAGCCACACCAGAAGAAACAAACGCGAGC
TTCAGTGGTGATTCCAGTGACTGGGGATAGTGAAGGTGCTACGGTGCAGCTGACTCCATA
TTTTCTACTTGTGGCAGCGACTGCATCCGACATAAAGGAACAGTTGTGCTCTGCCCACA
5 AACAGGCGTCCCTTTCCCTCTGGATAACAACAAAAGCAAGCCGGGANGGNCTGNCCTCTC
CTCCTGCTGTCTCTGCTGGTGGCCACATGGGTGCTGGTGGCAGGGATCTATCTAATGTGG
AGGCACGAAAGGATCAAGAAGACTTCCTTTTCTAACCACCACATTACTGCCCCCATTTA
AGGTTCTTGTGGTTTTACCCATCTGGAAATATGTTTTCCCTTCACACATTTGTTTATTTT
ATTGATTTNTTTCAAACCTTGGCAGGAGTTT

H25975

GGGTCCAGTGCAGTGGCTTGCNTGCAGAAAGAAGGCAGCAGACAAAGTCGTCTTCCTTCT
TTCCAATGACGTCAACAGTGTGTGCGATGGTACCTGTGGCAAGAGCGAGGGCAGTCCCAG
TGAGAACTCTCAAGACCTCTTCCCCCTTGCCTTTAACCTTTTCTGCAGTGATCTAAGAAG
15 CCAGATTCATCTGCACAAATACGTGGTGGTCTACTTTAGAGAGATTGATACAAAAGACGA
TTACAATGCTCTCAGTGTCTGCCCCAAGTACCACCTCATGAAGGATGCCACTGCTTTCTG
TGCAGAACTTCTCCATGTCAAGCAGCAGGTGTCAGCAGGAAAAAGATTCAAGCCTGCC
ACGATGGCTGCTTGCTTCCCTTTGTAGCCCACCCATGAGGAAGNCAAGAGACCTTNAAGG
GTTCCCTTTTCCCATCANTTTACAGGGGANAAAACGTGTGATGATC

H25941

TTTTGTTTGGCTNATNTNNTTCTTATTGTTGCATTATCAAAATGGTTATAGTTTTCAATT
AAAAGTGAATTGATTNCTATGTATAAAACAGCTTTGAAGTTGTAAATGTAGTTTCCAAT
CGTTAGTTAATGCTACATTAGTTAGCAATATTTGAAAATTTTATTGGTATAAAANGTTTT
25 AATTACTAAGGCTGTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACACGTTTTTCCC
CTGTAATTGGTGGGATAGGAAGCCTTTAAGGTCTCTNGCTTCTCATGGGTGGGCTACAAG
GAGCAGCAGCCATCGTGGCAGGCTTGTGANCTTTTNCCTGCTGACACCTGCTGCTTGACA
TGGGAGAAGTTCTGCACAGAAAGGCAGTGGGCATCCTTCATGAGGTGGGTACTTGGGGGN
CAGACACTGAGGAGCATTGT

BE539514

ACTCAAAGAAGGCAGCAGACAAAGTCGTCTTCCTTCTTTCCAATGACGTCAACAGTGTG
TGCGATGGTACCTGTGGCAAGAGCGAGGGCAGTCCCAGTGAGAACTCTCAAGACCTCTTC
CCCCTTGCCTTTAACCTTTTCTGCAGTGATCTAAGAAGCCAGATTACATCTGCACAAATAC
35 GTGGTGGTCTACTTTAGAGAGATTGATACAAAAGACGATTACAGTGCTCTCAGTGTCTGC
CCCAAGTACCACCTCATGAAGGATGCCACTGCTTTCTGTGCAGAACTTCTCCATGTCAAG
CAGCAGGTGTCAGCAGGAAAAAGATCACAAGCCTGCCACGATGGCCGCTGCTCCTTGTA
CCCACCCATGAGAAGCAAGAGACCTTAAAGGCTTCCTATCCCACCAATTACAGGGAAAAA
ACGTGTGATGATCCTGAAGCTTACTATGCAGCCTACAAACAGCCTTAGTAATTTAAACAT
40 TTTATACCAATAAAATTTTCAAATATGCTAACTAATGTAGCATTAACGATTGGAAA
CTACATTTACAACCTCAAAGCTGTTTTATACATAGAAATCAATTACAGCTTTAATTGAAA

ACTGTAACCATTTTGATAATGCAACAATAAAGCATCTTCAG

BX282554

5 GTCCAGTGCAGTGGCTTGCCACTCAAAAGAAGGCAGCAGACAAAGTCGTCTTCCTTCTTT
CCAATGACGTCAACAGTGTGTGCGATGGTACCTGTGGCAAGAGCGAGGGCAGTCCCAGTG
AGAACTCTCAAGACCTCTTCCCCCTTGCCCTTTAACCTTTTCTGCAGTGATCTAAGAAGCC
AGATTCATCTGCACAAATACGTGGTGGTCTACTTTAGAGAGATTGATACAAAAGACGATT
ACAGTGCTCTCAGTGTCTGCCCCAAGTACCACCTCATGAAGGATGCCACTGCTTTCTGTG
CAGAACTTCTCCATGTCAAGCAGCAGGTGTCAGCAGGAAAAAGATCACAAGCCTGCCACG
10 ATGGCCGCTGCTCCTTGTAGCCACCCATGAGAAGCAAGAGACCTTAAAGGCTTCCTATC
CCACCAATTACAGGGGAAAAAACGTGTGATGATCCTGAAGCTTACTAT

R74038

15 TATTGTTGCATTATCAAAATGGTTATAGTTTTCAATTAAACTGTAATTGATTTCTATGT
ATAAAACAGCTTTGAAGTTGTAAATGTAGTTTCCAATCGTTAGTTAATGCTACATTAGTT
AGCAATATTTGAAAATTTTATTGGTATAAAATGTTTTAATTACTAAGGCTGTTTGTAGGC
TGCATAGTAAGCTTCAGGATCATCACACGTTTTTNCCCTGTAATTGGGTGGGGATAGGGA
AGCCTTTAAGGTCTCTTGCTTCTCATGGGGTGGGGCTACAAGGGAGGCAGGCAGCCATCG
TGGGCAGGGCTTGTGATCTTTTTCCCTGCTGACACCTGCTGCTTGACATGGGGGGAAGGT
20 TCTGGCACAGAAAGCAGTGGGCATCCTTCATGAGGGTGGTACTTGGGGGGCAGACACTGA
GGAGGCNTTGTAATCGNCTTTTTTNGTATCCAANCTCTNCTAAAGTAGGGNCCACCNCGT
TTTTTNTTGCAGGTGGATNCGGGGCTN

R74129

25 GGGTCCAGTGCAGTGGCTTGCNTNCAAAAGAAGGCAGCAGACAAAGTCGTCTTCCTTCTT
TCCAATGACGTCAACAGTGTGTGCGATGGTACCTGTGGCAAGAGCGAGGGCAGTCCCAGT
GAGAACTCTCAAGACCTCTTCCCCCTTGCCCTTTAACCTTTTCTGCAGTGATCTAAGAAGC
CAGATTCATCTGCACAAATACGTGGTGGTCTACTTTAGAGAGATTGATACAAAAGACGAT
TACAATGCTCTCAGTGTCTGCCCCAAGTACCACCTCATGAAGGATGCCACTGCTTTCTGT
30 GCAGAACTTCTCCATGTCAAGCAGCAGGTGTCAGCAGGAAAAAGATCACAAGCCTGCCAC
GATNGCTGCTGCTCCTTGTAGNCCACCCATGAGAAGCAAGTGACCTTTAAAGGNTTTCCT
ATTNCCACCNATTTACAGGG

BG433769

35 GACTAGATGTTTGGCTGAAGATGCTTTATTGTTGCATTATCAAAATGGTTATAGTTTTCA
ATTAAAACAGCTTTGAAGTTGTAAATGTAGTTTCC
AATCGTTAGTTAATGCTACATTAGTTAGCAATATTTGAAAATTTTATTGGTATAAAATGT
TTTAATTACTAAGGCTGTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACACGTTTTT
TCCCTGTAATTGGTGGGATAGGAAGCCTTTAAGGTCTCTTGCTTCTCATGGGTGGGCTAC
40 AAGGAGCAGCAGCCATCGTGGCAGGCTTGTGATCTTTTTCTGCTGACACCTGCTGCTTG

ACATGGAGAAGTTCTGCACAGAAAGCAGTGGCATCCTTCATGAGGTGGTACTTGGGGCAG
ACACTGAGAGCATTGTAATCGTCTTTTGTATCAATCTCTCTAAAGTAGACCACCACGTAT
TTGTGCAGATGAATCTGGCTTCTTAGATCACTGCAGAAAAGGTTAAAGGCAAGGGGGAAG
AGGTCTTGAGAGTTCTCACTGGGACTGCCCTCGCTCTTGCCACAGGTACCATCGCACACA
5 CTGTTGACGTCATTGGAAAAGAAGGAAGAC

BG530489

GAGTTCTCACTGGGACTGCCCTCGCTCTTGCCACAGGTACCATCGCACACACTGTTGACG
TCATTGGAAAGAAGGAAGACGACCTTGTCTGCTACCTTCTTTTGAGTGGCAAGCCACTGC
10 ACTGGACCCATCTCTGCTATTTTCTTTTTCTGCCACTTTTCAAGGATGACCTCACTTCTG
CAATGGTTTTGAAGAAATTCAGTGAAGTAACAAATTGTGTGATGGAAACATATTTTCAGAT
GGGTAAACCACAAGAACCTTAATGGGGGGCAGTAGTGTGGTGGTAGAAAAGGAAGTCTTC
TTGATCCTTTCTGTGAGAGGAGAAAAGCATTTGTTATCTGTGAATAGCAAACAGCAGGCT
TTCCTCTGTAAACCATCCCTGACAAATGATCCCTTGCTAGAGAATGTCAGCTGAGCACC
15 AAGGGCCTTGTTAGTGACAGCAAGGAAAAACATCCTGATGTTCCTTTTGAACACATCACC
TGAAACACACTGATGCTTAAACCTTAACTTTTTTTTTTTGGGGGACATAGTCTCACTCTG
TCGCCCAGGCTGGAGTGCCTGGGAGAGGACCTCGGAAAGACTGGCAAGCATCCGCATACA
AGGGAGTAACAGCACAACTACTCCGTGAACTTCGGAGCCCTCCAAAGGAATACTCAAGGGC
GGGTAAAGGATGGCAAGGGTCGACGGAGAGCCACGAGGAGAGCGGAAGGTAGAGAGGAG
20 ACAAGCATAAGACGCGAGAGGAACTCCAAGGCGGGGCCAAAGAGAGAAACCACGGTCACC
AACAGAAG

AA007528

AGAAGCCAGATTTCATCTGCACAAATACGTGGTGNTCTACTTTAGAGAGATTGATACAAAA
25 GACGATTACAATGCTCTCAGTGTCTGCCCCAAGTACCACCTCATGAAGGATGCCACTGCT
TTCTGTGCAGAACTTCTCCATGTCAAGCAGCAGGTGTCAGCAGGAAAAAGATCACAAGCC
TGCCACGATGGCTGCTGCTCCTTGTAGCCCACCCATGAGAAGCAAGAGACCTTAAAGGCT
TCCTATCCCACCAATTACAGGGNAAAAACNGTAGTGATNATCCCTGACAGCTTACTATGC
CAGCCNT
30

AA007529

TTGGCTGAAGATGCTTTATTGTTGCATTATCAAAATCGGTTACAGTTTTCAATTAAAGCT
GTAATTNGATTTCTATGTATAAAACAGCTTTGAAGTTGTAAATGTAGTTTCCAATCGTTA
GTTAATGCTACATTAGTTAGCAATATTTGAAAATTTTATTGGTATAAAATGTTTTAATTA
35 CTAAGGCTGTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACACGTTTTTCCCTGTA
ATTGGGTGGGATAGGAAGCCTTTAAGGTCTCTTGCTTCTCATTGGGTGGGCTACAAGGAG
CAGCAGCCATCCGTNGGCAAGGCTTTGTGGATNCT

BI260259

40 GGAAGAGAAAGATCGTCCAGAGGTTCCATCGCACACACTGTATGACGTCATTGGAAATGA

AGGAAGACGACTTTGTCTGCTGGCTTCTTGTGAGTGGCAAGCCACTGCAGTGGACCCATC
TCTGCTATTTTCTTTATTCTGCCACTTTTCAAGGATGACCTCACTTCTGCAATGGTTTTG
AAGAAAGTTCAGTGAAGTAACAAATTGTGTGATGGAAACATATTTTCAGATGGGTAAACCA
CAAGAACCTTAATGGGGGGCAGTAGTGTGGTGGTAGAAAAGGAAGTCTTCTTGATCCTTT
5 CTGTGAGAGGAGAAAAGCATTAGTTATCTGTGAACAGCAAACAGCAGGCATTTTCACATCT
GTAAACCATCCCTGACAAATGATCCCTTGCTAGAGAATGTCAGCTGAGCACCAAGGGGCC
TTGTTAGTGACAGCAAGGACAAAACATCCTGATGTTCTTTTGAACACATCAGCTGAAAC
ACACTGATGCTCTAAACCGTTAACTATTTATTAATGGGGGAACATAGGTCTCAACTCATG
TACGACCAGGCTGGAGTGCAGTGGGGTTGAACATCGACAGACATAGCAAACCACCGATCA
10 CTAGGGAAACAACGCACAGAACTCCAGACTTAAAACACC

AA287951

ATTCGGCACCTGGGGGGCAGACACTGAGAGCATTGTAATCGTCTTTTGTATCAATCTCTC
TAAAGTAGACCACCACGTATTTGTGCAGATGAATCTGGCTTCTTAGATCACTGCAGAAAA
15 GGTAAAGGCAAGGGGGAAGAGGTCTTGAGAGTTCTCACTGGGACTGCCCTCGCTCTTGC
CACAGGTACCATCGCACACACTGTTGACGTCATTGGAAAGAAGGAAGACGACTTTGTCTG
CTGCCTTCTTTTGTAGTGGCAAGCCACTGCACTGGACCCATCTCTGCTATTTTCTTTTTCT
GCCACTTTTCAAGGATGACCTCACTTCTGCAATGGTTTTTGAAGAAATTCAGTGAAGTAAC
AAATNTGTGTGATGGAAACATATTTTCAGATGGGTAAACCACAAGAACCTTAATGGGGGGC
20 AGTAGTGTGGTGGTAGAAAAGGAAGTCTTCTTGATCCTTTCTGTGAGAGGAGAAAGC

AA287911

TTTTGATGGTCCACTTCCATTTAATGAATTAGTAAATATCTTTTCTCATGATTTTAATTA
CATTTTTTTTCTCTAGCTTACTTTATTATAATACAGCACATAATACACCTAACATGCAAAA
25 TATGTGTTAATTGGCTGTTTATGTTATTGGTAAGACTTCCAGTCAACAGTAGGCTATTAG
AAGTTAAGTTGTGGGAAAATCAAAGGTTATAGGAGATTTTCAACTGCATGCAGGGCCGGT
GCCCTCCCCACTGTGTTGTTCAAGGGTCAGCTGTACTCTCTAAGGGCTTTGCTAACTTCA
AAACATGGAGTATTTGAATACAGAAACCAGAGCATTTACATACTCAGCTCAAGGCAGAGC
TATTAAAAAAACTCCTCTTCTCCATATGTAGGAAAGGAAATACAAATGCATCCTTTGAGT
30 CATTTGTGATGT

T97852

AACAGTTGTGCTCTGCCCACAAACAGGCGTCCCTTTCCCTCTGGATAACAACAAAAGCAA
GCCGGGANGNCTGNCGCTCTCCTCCTGCTGTCTCTGCTGGTGGCCACATGGGTGCTGGTN
35 GCAGGGATCTATCTAATGTNGAGGCACGAAAGGGATCAAGAGGACTTCCTTTTCTACCAC
CACACTACTGCCCCCATTAAGGTTCTTGNGGTTTACCCATCTGGAAATATGTTTCCAT
CACACAATTTGTTACTTCACTGGAATTTCTTCAAACCATTTGGCAGGANGTGAGGGTCAT
CCTTGAAAAGTGGGC

40 T97745

CCTCACTTCTGCAATGGTTTTGAAGAAATTCAGTGAAGTAACAAATTGTGTGATGGAAAC
ATATTTTCAGATGGGTAAACCACAAGAACCTTAATGGGGGGCAGTAGTGTGGTGGTAGAAA
AGGAAGTCTTCTTGATCCTTTTCGTGCCTCCACATTAGATAGATCCCTGCCACCAGCACCC
ATGTGGCCACCAGCAGAGACAGCAGGAGGAGAGGCAGCCAGCCTCCCGGCTTTGCTTTTG
5 TTGTTATCCAGAGGGGAAAGGGGACGCCTGTTTNTGGGGCAGAGCACAACTGTTTCCCTC
GTGCCCCGAATTCTTTGGGCCTTCGAGGGGCCAAATTTCCCTATTAGGTGAGGTCGTATTT
TAAATTTTCGGTAATTCATGGTCATAGGCTTGTTTTTCCCCG

N40294

10 GTTTCAACACAATTTTGGATCAGCTGCCTGTTTGCAAAAACATAATATATTTCTGTTAAA
CAGTTCCTTCACCTAACAGCATATTGCTCTTATAACTGGTAGAGCTGTTTCAAAGGAAGTT
GGTTTCTGGTCCAAGTTTTGACCTAAACCATGTCCATCTTCTATTACCAGCACTTACAAG
CACTGTGAAAACCTGATCATGACAAATAAGTAAATTTGCTACATTAAACATATTGCCTCA
GCCATTACTAAGCGTCCACTTGTAAGCTGGACACAGTTTTTACTTTTATGCTTCATTTTG
15 ATTTTTTATCCGTAAGACATAAATTAGAAGGCATGAGGTGGCCCTTTAAGGATAATCTGC
AAATATACACATTTTAAATAGTCATCCATCTGGAAATCGNTCCACCATTCCAGGGGAAGG
ATTCCAGGTATTGGTGCTGTGGTGGAATAAAGCATTCCCCNGGGAAAAAAACCATTTTA
TGNCTAAATAATTACCACCATTAACTCNTGGGGTT

20 AA809841

GAATACTAACTCTAAGAACCCCTCACTGATTCCTCAATAGCATCTTAAGTGAAAAACCT
TCTATTACATGCAAAAAATCATTGTTTTTAAGATAACAAAAGTAGGGAATAACAAGCTG
AACCCTACTTTTACTGGACCAAATGATCTATTATATGTGTAACCACTTGTATGATTTGGGA
TTTGCAT

25

AA832389

TTTTTTACAACCTCAAAGCTGTTTTATACATAGAAATCAATTACAGTTTTTAATTGAAAAC
TATAACCATTTTGATAATGCAACAATAAAGCATCTTCAGCCAAACATCTAGTCTTCCATA
GACCATGCATTGCAGTGTAACCCAGAACTGTTTAGCT

30

H14692

CTGAGTGTGATGGTGTAAGCCTGTGGTCCCAGCTACTAGGGAGGCTGAGATGGGATTACA
GGTGTGAGCCACGGCGCCTGGCCTAAAAGCATCTTTTTCTTTAACGCAGAGGTTATGTTG
TATTATTAGCATAAATGTTTTTTCTGGGAATGCTTATTTACACAGCACAACTGAAT
35 CTTCTCTGGAATGTGGATCGATTTTCAGATGGATGACTATTAAATGTGTATATTTGCAGA
TTATCCTTAAAGGGCCACCTCATGCCTTCTAATTTATGTCTTACGGATAAAAAATCAAAA
TGAAGCATAAAGTAAAACTGTGTCCAGCTTTACAAGTGACGCTTAGTAATGGCTGAGG
CAATATGTTTAAATGTAGCCAAATTTTACTTATTTGTCCATGATCCAGTTTTTTCAGAGTGC
TTGTTAAGTGCTGGTAATTAGGAAGGTGGGACATGGGTAGGTCAAACTTGGGACCNGA
40 AACCAACTTGN

AA732635

TTTTTTTTTTTACAACCTTCAAAGCTGTTTTATACATAGAAATCAATTACAGTTTTAATTGA
AAACTATAACCATTTTGTAAATGCAACAATAAAGCATCTTCAGCCAAACATCTAGTCTTC
5 CATAGACCATGCATTGCATTGTACCCAGAACTGTTTAGCTAATATTCTATGTTTAATTAA
TGAATACTAACTCTAAGAACCCCTCACTGATTCACTCAATAGCATCTTAAGTGAAAAACC
TTCTATTACATGCAAAAAATCATTGGTTTT

AA928257

10 TTTTCTGAGTAAGAACAGGCTTTATTTGTAAAACCACTCGTGACTCTTTACAAAGCAGGA
TACACAGAAGGGAAAAAATACACAGTGCAAAATGGATGTTCTGAGTGCCACAAGGATCT
GCTGAAAAAAGCCAAAGATGTAAGATGGCTGGGTATATATGAGAATGAATATTTCACTAT
ATTCTGATTCAATTACCAGTCTCAGTGGCCCAGGATGAGCTTTTGGTGTGGTCACATGGC
CAACATTTGGATAACAAATGAGGAATAATGGTACCGCCTCACTAGTGCCTGAGAACAGCA
15 TGTTCTGGAAAATGTCTCTGGAGTTAGAGATGTGTTAGCTTTTTTCATTACAGATGGAGAA
ATACAATGTTTACACAACAGTCCAGGGGTGGGGTCAAAAGTTGGAAGGTGTCATTAGACG
CAGCCAAATAAAGTGAAGACAACCCAGGTGACTGGCAGCCCTGACTTGTGCGTGGGCG

AI184427

20 TTTCTGAGTAAGAACAGGCTTTATTTGTAAAACCACTCGTGACTCTTTACAAAGCAGGAT
ACACAGAAGGGAAAAAATACACAGTGCAAAATGGATGTTCTGAGTGCCACAAGGATCTG
CTGAAAAAAGCCAAAGATGTAAGATGGCTGGGTATATATGAGAATGAATATTTCACTATA
TTCTGATTCAATTACCAGTCTCAGTGGCCCAGGATGAGCTTTGGTGGTGGTCACATGGCC
AACATTTGGATAACAAATGAGGA

25

AI298577

GAGATGGAGGTCTCGCTTTGTGACGTAGCCTGGTCTTGAGCGATCCTTTTGCCTTGGCCT
TGCCAAAGTGCTGGGATTGGAGGCATGAGCCACTGCACCCACCCCTGTTTTTTTTTTAAG
TAAACCATTATAATAACTCATTTATAAAAAGGTTACTTCAAGAGGGCTTTCAACTTAAGA
30 ATTATTTTCATTTTGAACATGAAAAGTTAAATAGTAACTAAGAACTGAGAACTCTGACA
GTGACCTCTAATAGGTAACCTTAGGCAAAAGTAGACAAGTTTGTGGGTATTTTGTGTGTTT
ATGTTAAAAGGCACCTGTACAAGAATCAAGATATGAATCTAGTTTGTAGAGGGAAGGTCT
TATGCAAATACCAAATCATAAAGTGGT

35 AI692717

AGAGATGTTGGTCTCGCTTTGTGACGTAGCCTGGGCTTGAGCGATCCTTTTGCCTTGGCC
TTGCCAAAGTGCTGGGATTGGAGGCATGAGCCACTGCACCCACCCCTGTTTTTTTTTTTAA
GTAAACCATTATAATAACTCATTTATAAAAAGGTTACTTCAAGAGGGCTTTCAACTTAAG
AATTATTTTCATTTTGAACATGAAAAGTTAAATAGTAACTAAGAACTGAGAACTCTGAC

AGTGACCTCTAATAGGTAACCTTTAGGCCAAAAGTAGACAAGTTTGTGGGTATTTTGTGTGTT
CATGTTAAAAGGCACCTGTACAAGAATCAAGATATGAATCTAGTTTGTAGAGGGAAGGTC
TTATGCCAAATACCAAATCATACAAGTGGTTACACATATAATAGATCATTGGTCCAGTAA
AAGTGGGTTCAGCTTGTTTATTCCTACTT

5

AA910922

GAGATGGAGGTCTCGCTTTGTGACGTAGCCTGGTCTTGAGCGATCCTTTTGCCTTGGCTT
GCAAAGTGCTGGGATTGGAGGCATGAGCACTGCACCCACCCCTGTTTTTTTTTTTAAAGTA
AACCATTATAATAACTCATTTATAAAAAGGTTACTTCAAGAG

10

H90761

TTCACTCAATAGCATCTTAAGTGAAAAACCTTCTATTACATGCAAAAAATCATTGTTTTT
AAGATAACAAAAGTAGGGAATAAACAAGCTGAACCCACTTTTACTGGACCAAATGANCTA
TTATATGTATAACCACTTGTATGATTTGGTATTTGCATAAGACCTTCCCTCTACAACTA
GATTCATATCTTGATTCTTGTACAGGTGCCTTTTTTAATATTCTGTGATGAAATCGTTCAC
AGTCAGAGTACATGTCTGCTGCATATGGGAAATAGGGACTGTTGTTCTGAGGGACAAGGC
ACTCAATTCAGCCGTAAAGGCTGACCCGGGCTACTTTTTTTCCANGGGAATACAATTTTT
TTACCTTGAATAAAAATNGGGCCCGACNGGAC

15

20

AI620122

TTTTTTTTTTTGAGTAAGAACAGGCTTTATTTGTAAAACCACTCGTGACTCTTTACAAAG
CAGGATACACAGAAGGGAAAAAATACACAGTGCAAAATGGATGTTCTGAGTGCCACAAG
GATCTGCTGAAAAAAGCCAAAGATGTAAGATGGCTGGGTATATATGAGAATGAATATTT
CACTATATTCTGATTCAATTACCAGTCTCAGTGGCCCAGGATGAGCTTTTGGTGTGGTCA
CATGGCCAACATTTGGATAACAAATGAGGAATAATGGTACCGCCTCACTAGTGCCTGAGA
ACAGCATGTTCTGGAAAATGTCTCTGGAGTTAGAGATGTGTTAGCTTTTTTCATTACAGAT
GGAGAAATACAATGTTTACACAACAGTCCAGGGGTGGGGTCAAAGTTGGAAGGTGTCAT
TAGACGCA

25

30

AI793318

AAATTTTTAACTTTTAATAGTTAAAATAGTTAACTATTGGTATGGTAGGAAATGATAAAG
TAGACTAGTATCTGTATACATTTTCTGCATTTATGACATACCTTTTTCTTCATTTTTTTC
AATATTTTAATTGAAAAGTTCATCCGAGTTTCATCTAAGTTTTTTCAAAGTGATACAAAT
CTCCAAAAAATTTTCCAATATATGTATTGAAAAAATCCAGGTGTAAGTGGCTCTGCGCAG
TCCAAACCTGTGTTGTTCAAGGGTCAACTGTGTATGAATCCAAGCGAAAGCTTTTCTTAA
CACCTCATAAGAACTATTTTTTAAAAAACAGGAACTAGCATAGAGTAACCATCACAGGTA
AAGTGTAATTTGTTATCAGCCATCTTTGCCCATTTCACTACTGGTAGAAGGCTCAATGG
TAAAAATAAA

35

40

AA962325

TTTTTTTTTTTTTTTTTTTTTTTTCTGACTGTCCCGTTTTTATTTTTACCATTGAGCCTT
CTACCAGTACTGAAATGGGCAAAGATGGCTGATAACAAATTACACTTTACCTGTGATGG
TTACTCTATGCTAGTTCCTGTTTTTTAAAAAATAGTTCTTATGAGGTGTTAAGAAAAGCT
TTCGCTTGGATTTCATACACAGTTGACCCTTGAACAACACAGGTTTGGACTGCGCAGACCA
5 CTTACACCTGGATTTTTTCAATACATATATTGGAAAATTTTTTGGGGATTTGTATCACTT
TGAAAAAACTTAGATGAAACTCGGATGGACTTTTCCATTAAAATATTGGAAAAAATGAAG
AAAAAGGT

AI733290

10 TTTTTTTTTTTTTTTTTTTTTTTTTCTGACTGGCCCGTTTTTATTTTTACCATTGAGCCTT
CTACCAGTACTGAAATGGGCAAAGATGGCTGATAACAAATTACACTTTACCTGGGATGG
TTACTCTATGCTAGTTCCTGTTTTTTAAAAAATAGTTCTTATGAGGGGTAAAAAAGCT
TTCGCTTGGATTTCATACACAGTTGACCCTTGAACAACACAGGTTTGGACTGCGCAGAGCC
ACTTACACCTGGATTTTTTCAATACATATATTGGAAAATTTTTTGGAGATTTGTATCACT
15 TTGAAAAAACTTAGATGAAACTCGGATGAACTTTTCAATTAAAATATTGAAAAAATGAA
GAAAAAGGTATGTCATAAATGCAGAAAATGTATACAGATACTAGTCTACTTTATCATTTT
CTACCATACCAATAG

BQ226353

20 TAAAGGAACAGTTGTGCTCTGCCCACAAACAGGCGTCCCTTTCCCTCTGGATAACAGTAA
GTGCCCAGTAACTTCAACCAGATGATCAAAGTGGCTCACACACAGTCACTGCCCCCCT
CAGTATGTGGAAGGGTTGTGTGTATGTGGGCAGTGCAAGGGGTGCTGCCTGTGTACT
GAACTGGGGTGCAGAGAAAGCCAACAGTGCTGTCCAGAGAACCTAGAATCTGAGTAAGA
ACAGGCTTTATTTGTAAACCCTCGTGACTCTTTACAAAGCAGGATACACAGAAGGGAA
25 AAAAATACACAGTGCAAAATGGATGTTCTGAGTGCCACAAGGATCTGCTGAAAAAGCCA
AAGATGTAAGATGGCTGGGTATATATGAGAATGAATATTTCACTATATTCTGATTCAATT
ACCAGTCTCAGTGGCCCAGGATGAGCTTTTGGTGTGGTCACATGGCCAACATTTGGATAA
CAAATGAGGAATAATGGTACCGCCTCACTAGTGCCTGAGAACAGCATGTTCTGGAAAATG
TCTCTGGAGTTAGAGATGTGTTAGCTTTTTTCATTACAGATGGAGAAATACAATGTTTACA
30 CAACAGTCCAGGGGTGGGGTCAAAAGTTGGAAGGTGTCATTAGACGCAGCCAAATAAAG
TGAAGACCACCCAGGTGACTGGCAGCCCTGACTTGTGCGTGGGCGAAACCTTACAGATTC
CTGGGGCACTCTGTGCCTGAACTTACCTGGATGGTCTTTGTGAGGCGGGTGGGCACTTAT
CCTCCATNAATGGTCAGTCTAACAAGACCGGCCTGTAAAAATGGCATCTAATAGGGGCTA
TGAATGGAACAGTTGGTACCCAGAAATAACTTTAATT
35

W04890

GACAGTCTGGGAGCCCAGAGCTCTGGGAGGAGTNGGGAAAATGCTGCTTCCTGCTGCTTG
CTTCTAGGCACCTGCTTCCGCCATCTCACTTACCATGGCTAGAGATGGGGGTGAGACTGG
GGAAGGACAAAAGCAGGGAACAGATAAGGGATGGAAATCAGAAGGGAATATAGAAAGAAC
40 TCTGGATATGCNAGAAATGCCGGTACCTGAGCATTTTGTATCAATGGGAGTACCCTCTGT
AACTGCTCAGTAGGTTACAAATGAAGAGTCCACCAGTATTAGAAACAATTTAACTTGCC

AGTACCAACTGGGATGTGTGCCTTCAATTTGAAAATTGTATGTTTTATTTTTTAAATTTG
GTTAACAGCATTAAATTTATAGAGTATTTGATGTCATTTATGGTTCCCGAGGTGTTTCCAA
CACAATTTTTGGGATCA

5 BM455231

CTTTTAATAGTTAAAATAGTTAACTATTGGTATGGTAGGAAATGATAAAGTAGACTAGTA
TCTGTATACATTTTCTGCATTTATGACATACCTTTTTCTTCATTTTTTTTCAATATTTTAA
TTGAAAAGTTCATCCGAGTTTCATCTAAGTTTTTTCAAAGTGATACAAATCTCCAAAAAA
TTTTCCAATATATGTATTGAAAAAATCCAGGTGTAAGTGGCTCTGCGCAGTCCAAACCTG
10 TGTTGTTCAAGGGTCAACTGTGTATGAATCCAAGCGAAAGCTTTTCTTAACACCTCATAA
GAACTATTTTTTAAAAAACAGGAACTAGCATAGAGTAACCATCACAGGTAAAGTGTAATT
TGTTATCAGCCATCTTTTGCCCATTTTCAGTACTGGTAGAAGGCTCAATGGTAAAAATAAA
AACGGGACAGTCAGAAGATCTGGAAGTCCTGACCCTGCTTTCACCTGGCATGTGTAATCC
AGTCATGCTCGTATCAGTCTCTGTAGGAGCACTTGAAGGTATTACATAAATGCTATCTAA
15 CTCTGGGAAACGCCAACATGTGATTGCCTCCAGAGGAATCTTCTTTAAAAAAAATTCAA
AATGTTATTTCTTACTAGGATGTCTTTAAAGAATTATAACCCTTACCGTGCCTCCACAT
TAGATAGATCCCTGCCACCAGCACCCATGTGGCCACCAGCAGAGACAGCAGGAGGAGAGG
CAGCCAGCCTCCCGGCTTGCTTTTGTCTGGAAAAAACAAAGCTTATTCACCTTTGGAAA
AAAATCCACACTTATCTCTTAATTTAAAAACTAAGACTTGGTATACTTTATAGAGGGTTA
20 TTTATTTTTTATTATTTTTTAGTTTTGAGACAGAGTCTCGCTTTGTTGCCTANGCTGGAG
TGCAGTGGCGCAATCTCGGTTCACTGCAGCCTCCGTTCTCCGGGGTTCAAGGCATGCTGG
CTCAGCCTCCTGTATAGCTGGGGATTAAAGGCATGTGTTACGCGGCCCCAGCCCCTTTTG
TAAAAGATTTAGATCCCTTTTAAAACCATCAGTCAGGAGGCTCCTTTAAAAGTCTGGCC
ATCTAATCTTTTTTCCCCCAAAGGGG

25

BI492426

TTTTTTTTTTTTCTTTTTCTGAGTAAGAACAGGCTTTATTTGTAAAACCACTCGTGACTC
TTTACAAAGCAGGATACACAGAAGGGAAAAAATACACAGTGCAAAATGGATGTTCTGAG
TGCCACAAGGATCTGCTGAAAAAGCCAAAGATGTAAGATGGCTGGGTATATATGAGAAT
30 GAATATTTCACTATATTCTGATTCAATTACCAGTCTCAGTGGCCCAGGATGAGCTTTTGG
TGTGGTCACATGGCCAACATTTGGATAACAAATGAGGAATAATCTCGTGC

BG674622

AATTTATAGAGTATTGATGTCATTTATGTTTCTGAGGTGTTTCAACACAATTTTGGATCA
35 GCTGCCTGTTTGCAAAAACATAATATATTTCTGTAAACAGTTCTTCACCTAACAGCATA
TTGCTCTTATAACTGGTAGAGCTGTTTCAAAGGAAGTTGGTTTCTGGTCCAAGTTTTGAC
CTAAACCATGTCCATCTTCTATTACCAGCACTTACAAGCACTGTGAAAACCTGATCATGAC
AAATAAGTAAAATTTGCTACATTAAACATATTGCCTCAGCCATTACTAAGCGTCCACTTG
TAAAGCTGGACACAGTTTTTACTTTATGCTTCATTTTGATTTTTTATCCGTAAGACATAA
40 ATTAGAAGGCATGAGGTGGCCCTTTAAGGATAATCTGCAAAATATACACATTTTAAATAGTC
ATCCATCTGAAATCGATCCACATTCCAGAGAAGATTAGTATTGTGCTGTGTGAAATAAG

CATTCCCAGAAAAAAACATTTATGCTAATAATACAACATAACCTCTGCATTAAAGAAAA
AGATGCTTTTAGGCCAGGCGCCGTGGCTCACGCCTGTAATCCCTGCACTTTGAGAGGCTG
AGGTGGGTGGATCATGAGGTGAGGAGATCAAGACCATCCTGGCTAACAGGGTGAAACCCC
GTCTCTACTGGGGATATAACAAAGTTAGCTGGGTGTGGTGGTGGGTGCTTGTGGTCCCAG
5 CTACTCAGGAGGCTGAGGCAGGAGAATGGCGTGAACCCGGAAGGCAGAGGTTGTAGTGAC
GCGAGGTTACGCCACTGCATTCCAGTCTGGG

BX111256

10 CAGGAAGNTAAGAACAGTCCTAAAATCTCTTTGGCTTCTTTGTCCTGATATGCACCGGCA
TTTTACAGTAGGAAGTAGGGTTTCTGTCCAGTTTTTTTTGGTTCTTTAAGGAATTAATGT
TATTCTGGGTACAACCTGCTTACATACATAGCACATATAGATGACATTTTTACAGGCCGTC
TTGTTAGACTGACATACATGGAGGATAGTGCCACCCGCCTCACAAGAACATCAGGTAAGC
TCAGGCACAGAGTGCCCAGGAATCTGTAAGGCTTCGCCCACGCACAAGTCAGGGCTGCCA
15 GTCACCTGGGTGTCTTCACTTTATTTGGCTGCGTCTAATGACACCTTCCAACCTTTTGAC
CCCACCCCTGGACTGTTGTGTAAACATTGTATTTCTCCATCTGTAATGAAAAAGCTAACA
CATCTCTAACTCCAGAGACATTTTCCAGAACATGCTGTTCTCAGGCACTAGTGAGGCGGT
ACCATTATTCCTCATTTGTTATCCAAATGTTGGCCATGTGACCACACCAAAGCTCATCC
TGGGCCACTGAGACTGGTAATTGAATCAGAATATAGTGAAATATTCAATTCTCATATATAC
20 CCAGCCATCTTACATCTTTGGCTTTTTTTCAGCAGATCCTTGTGGCACTCAGAACATCCAT
TTTGCACTGTGTATTTTTT

BX117618

25 AAATTTTTTAACTTTTTAATAGTTAAAATAGTTAACTATTGGTATGGTAGGAAATGATAAAG
TAGACTAGTATCTGTATACATTTTCTGCATTTATGACATACCTTTTTCTTCATTTTTTTC
AATATTTTAATTGAAAAGTTCATCCGAGTTTCATCTAAGTTTTTCAAAGTGATACAAAT
CTCCAAAAAATTTTCCAATATATGTATTGAAAAAATCCAGGTGTAAGTGGCTCTGCGCAG
TCCAAACCTGTGTTGTTCAAGGGTCAACTGTGTATGAATCCAAGCGAAAGCTTTTCTTAA
CACCTCATAAGAACTATTTTTTAAAAAACAGGAACTAGCATAGAGTAACCATCACAGGTA
AAGTGTAATTTGTTATCAGCCATCTTTTGCCCATTTCACTACTGGTAGAAGGCTCAATGG
30 TAAAAATAAAAACGGGACAGTCAGAAAAA

AA682806

35 TCTGAGTAAGAACAGGCTTTATTTGTAAAACCACTCGTGACTCTTTACAAAGCAGGATAC
ACAGAAGGGAAAAAATACACAGTGCAAAATGGATGTTCTGAGTGCCACAAGGATCTGCT
GAAAAAAGCCAAAGATGTAAGATGGCTGGGTATATATGAGAATGAATATTTCACTATATT
CTGATTCAATTACCAGTCTCAGTGGCCAGGATGAGCTTTTGGTGTGGTCACATGGCCAA
CATTTGGATAACAAATGAGGAATAATGGTACCGCCTCACTAGTGCCTGAGAACAGCATGT
TCTGGAAAATGTCTCTGGAGTTAGAGATGTGTTAGCTTTTTTCATTACAGATGGAGAAATA
CAATGTTTACACAACAGTCCAGGGGTGGGGTCAAAG
40

AI202376

CTGACTGTCCCGTTTTTATTTTTACCATTGAGCCTTCTACCAGTACTGAAATGGGCAAAA
GATGGCTGATAACAAATTACACTTTACCTGTGATGGTTACTCTATGCTAGTTCCTGTTTT
TTAAAAAATAGTTCTTATGAGGTGTTAAGAAAAGCTTTCGCTTGGATTACATACACAGTTG
ACCCTTGAACAACACAGGTTTGGACTGCGCAGAGCCACCCTCGTGCCGAATT

5

AI658949

CTGACTGTCCCGTTTTTATTTTTACCATTGAGCCTTCTACCAGTACTGAAATGGGCAAAA
GATGGCTGATAACAAATTACACTTTACCTGTGATGGTTACTCTATGCTAGTTCCTGTTTT
TTAAAAAATAGTTCTTATGAGGTGTTAAGAAAAGCTTTCGCTTGGATTACATACACAGTTG
ACCCT

10

BG403405

GGAAATGATAAAGTAGACTAGTATCTGTATACATTTTCTGCATTTATGACATACCTTTTT
CTTCATTTTTTTCAATATTTTAATTGAAAAGTTCATCCGAGTTTCATCTAAGTTTTTTCA
AAGTGATACAAATCTCCAAAAAATTTTCCAATATATGTATTGAAAAAATCCAGGTGTAAG
TGGCTCTGCGCAGTCCAAACCTGTGTTGTTCAAGGGTCAACTGTGTATGAATCCAAGCGA
AAGCTTTTCTTAACACCTCATAAGAACTATTTTTTAAAAAACAGGAAGTAGCATAGAGTA
ACCATCACAGGTAAAGTGTAATTTGTTATCAGCCATCTTTGCCCATTTTCAGTACTGGTAG
AAGGCTCAATGGTAAAAATAAAAAACGGGACAGTCAGAAGATCTGGAAGTCCTGACCCTGC
TTTCACCTGGCATGTGTAATCCAGTCATGCTCGTATCAGTCTCTGTAGGAGCACTTGAAG
GTATTACATAAATGCTATCTAACTCTGGGAAACGCCAACATGTGATTGCCTCCAGAGGAA
TCTTCTTTAAAAAAAATTCAAAATGTTATTTCCCTTACTAGGATGTCTTTAAAGAATTAT
AACCTTACCCTGCCTCCACATTAGATAGATCCCTGCAACAGACCCATGTGGCACCAGCA
GAGACAGCAGGAGGAGAGGCAGCAGCTCCCGGTTGTTTGTCTGGAAAAACAAAGGTTATC
ACTTTG

15

20

25

BE673417

CTGACTGTCCCGTTTTTATTTTTACCATTGAGCCTTCTACCAGTACTGAAATGGGCAAAA
GATGGCTGATAACAAATTACACTTTACCTGTGATGGTTACTCTATGCTAGTTCCTGTTTT
TTAAAAAATAGTTCTTATGAGGTGTTAAGAAAAGCTTTCGCTTGGATTACATACACAGTTG
ACCCT

30

AW021469

GCACGAGATTATTCCTCATTTGTTATCCAAATGTTGGCCATGTGACCACACCAAAAAGCTC
ATCCTGGGCCACTGAGACTGGTAATTGAATCAGAATATAGTGAAATATTCATTCTCATAT
ATACCCAGCCATCTTACATCTTTGGCTTTTTTCAGCAGATCCTTGTGGCACTCAGAACAT
CCATTTTGCAGTGTGTATTTTTTTCCCTTCTGTGTATCCTGCTTTGTAAAGAGTCACGAG
TGTTTTTACAAATAAAGCCTGTTCTTACTCAGAAAAAAAAAAAAAAAAAAAAA

35

40

CF455736

NNTTGAACAGGCGTGACGGTCCGGATTCCCGGGATGTTGTGCTCTGCCCACAAACAGGCG
TCCCTTTCCCTCTGGATAACAACAAAAGCAAGCCGGGAGGCTGGCTGCCTCTCCTCCTGC
TGTCTCTGCTGGTGGCCACATGGGTGCTGGTGGCAGGGATCTATCTAATGTGGAGGCACG
AAAGGATCAAGAAGACTTCCTTTTCTACCACCACACTACTGCCCCCATTAAAGGTTCTTG
5 TGGTTTACCCATCTGAAATATGTTTCCATCACACAATTTGTTACTTCACTGAATTTCTTC
AAAACCATTGCAGAAAGTGAGGTCATCCTTGAAAAGTGGCAGAAAAAGAAAATAGCAGAGA
TGGGTCCAGTGCAGTGGCTTGCCACTCAAAAGAAGGCAGCAGACAAAGTCGTCTTCCTTC
TTTCCAATGACGTCAACAGTGTGTGCGATGGTACCTGTGGCAAGAGCGAGGGCAGTCCCA
GTGAGAACTCTCAAGACCTCTTCCCCCTTGCCCTTTAACCTTTTCTGCAGTGATCTAAGAA
10 GCCAGATTCTCATCTGCACAAATACGTGGTGGTCTACTTTAGAGAGATTGATACAAAAGACG
ATTACAATGCTCTCAGTGTCTGCCCCAAGTACCACCTCATGAAGGATGCCACTGCTTTCT
GTGCAGAACTTCTCCATGTCAAGCAGCAGGTGTCAGCAGGAAAAAGATCACAAGCCTGCC
ACGATGGCTGCTGCTCCTTGCTAGCCCACCCATGAGAAGCAAGAGACCTTNAAGGCTTCCT
ATCCCACCATTACAG

AW339874

TTTTTTTTTTTTTCTGAGTAAGAACAGGCTTTATTTGTAAAACCACTCGTGACTCTTTAC
AAAGCAGGATACACAGAAGGGAAAAAATACACAGGGCAAAATGGATGTTCTGAGTGCCA
CAAGGATCTGCTGAAAAAGCCAAAGATGTAAGATGGCTGGGTATATATGAGAATGAATA
20 TTTCACTATATTCTGATTCAATTACCAGTCTCAGTGGCCAGGATGAGCTTTTGGTGTGG
TCACATGGCCAACATTTGGATAACAAATGAGGAATAATGGTACCGCCTCACTAGTGCCTG
AGAACAGCATGTTCTGGAAAATGTCTCTGGAGTTAGAGATGTGTTAGCTTTTTCATTACA
GATGGAGAAATACAATGTTTACACAAC

BG399724

CATGATGTTCAGTATGATCAGTTAACCTTAACCTCTGAGCATCCTGAAGCAAAATCTAAA
TAATGCAGCTATTACCACTGGTGGTCCAGGCTCTGGTGAAGCCCTCTGAGCCCAGGAGGA
AGAGAAAGCATTGTCCAGAGGTAGGAACACAGTCTGGGAGCCCAGAGCTCTGGGAGGAGT
GGGAAAATGCTGCTTCCTGCTGCTTGCTTCTAGGCACCTGCTTCCGCCATCTCACTTACC
30 ATGGCTAGAGATGGGGGTGAGACTGGGGAAGGACAAAAGCAGGGAACAGATAAGGGATGG
AAATCAGAAGGGAATATAGAAAGAACTCTGGATGTGGAGAAATGCCGGTACCTGAGCATT
TTGTATCAATGGGAGTACCCTCTGTAACCTGCTCAGTAGGTACAAATGAAGAGTCCACCA
GTATTAGAAACAATTTAAACTTGCCAGTACCAACTGGGATGTGTGCCTTCAATTTGAAAA
TTGTATGTTTTATTTTTTAAATTTGTTAACAGCATTAATTTATAGAGTATTGATGTCATT
35 TATGTTTCTGAGGTGTTTCAA

BF475787

TCTGAGTAAGAACAGGCTTTATTTGTAAAACCACTCGTGACTCTTTACAAAGCAGGATAC
ACAGAAGGGAAAAAATACACAGTGCAAAATGGATGTTCTGAGTGCCACAAGGATCTGCT
40 GAAAAAAGCCAAAGATGTAAGATGGCTGGGTATATATGAGAATGAATATTTCACTATATT
CTGATTCAATTACCAGTCTCAGTGGCCAGGATGAGCTTTTGGTGTGGTCACATGGCCAA

CATTTGGATAACAAATGAGGAATAATGGTACCGCCTCACTAGTGCCTGAGAACAGCATGT
TCTGGAAAATGTCTCTGGAGTTAGAGATGTGTTAGCTTTTTTCATTACAGATGGAGAAATA
CAATGTTTACACAACAGTCCAGGGGTGGGGTCAAAAGTTGGAAGGTGTCATTAGACGCAG
CCAAATAAAGTGAAGACAACCCAGGTGACTGGCAGCCCTGACTTGTGCGTGGGCGA

5

BF437145

CTGACTGTCCCGTTTTTATTTTTTACCATTGAGCCTTCTACCAGTACTGAAATGGGCAAAA
GATGGCTGATAACAAATTACACTTTACCTGTGATGGTTACTCTATGCTAGTATCCTGTTT
TTTAAAAAATAGTTCTTATGAGGTGTTAAGAAAAGCTTTCGCTTGGATTCATACACAGTT
GACCCT

10

H64601

AGGAAGTTAAGAACAGTCCTAAATCTCTTTGGCTTCTTTGTCCTGATATGCACCGGCAT
TTTCACAGTAGGAACTAGGGTTTCTGTCCAGTTTTTTTTGGTTCTTTAAGGAATTAATGTT
ATTCTGGGTACAACGCTTACATACATAGCACATATAGATGACATTTTTTACAGGCCGTCT
TGTTAGACTGACATACATGGAGGATAGTGCCACCCGCCTCACAAGAACATCAGGTAAGCT
CAGGCACAGAGTCCNAGGGNATCTGTAAGGGCTTCGCCCACGCACAAGTCAGGGCTGCCA
GTCACCNGGGTTGTCTTCACTTTATTTGGGCTGCGTCTAATGACACCTTNCCAACTTTTT
GACCCACCCCTGGGGCTTGTTGTGTAAACCATTGTTATTTCTCCCNTCTGTAATGGAAAA
AGGTTAACACNTTTTTAACTTCCGGNGACATTTTTTC

15

20

AF212365

gcacgagcga	tgtcgctcgt	gctgctaagc	ctggccgcgc	tgtgcaggag	cgccgtaccc
cgagagccga	ccgttcaatg	tggctctgaa	actgggccat	ctccagagtg	gatgctacaa
catgatctaa	tccccggaga	cttgagggac	ctccgagtag	aacctgttac	aactagtgtt
gcaacagggg	actattcaat	tttgatgaat	gtaagctggg	tactccgggc	agatgccagc
atccgcttgt	tgaaggccac	caagatttgt	gtgacgggca	aaagcaactt	ccagtcctac
agctgtgtga	ggtgcaatta	cacagaggcc	ttccagactc	agaccagacc	ctctggtggt
aaatggacat	tttcctacat	cggcttcctt	gtagagctga	acacagtcta	tttcattggg
gcccataata	ttcctaatac	aaatatgaat	gaagatggcc	cttccatgtc	tgtgaatttc
acctcaccag	gctgcctaga	ccacataatg	aaatataaaa	aaaagtgtgt	caaggccgga
agcctgtggg	atccgaacat	cactgcttgt	aagaagaatg	aggagacagt	agaagtgaac
ttcacaacca	ctcccctggg	aaacagatac	atggctctta	tccaacacag	cactatcatc
gggttttctc	aggtgtttga	gccacaccag	aagaaacaaa	cgcgagcttc	agtggtgatt
ccagtgactg	gggatagtga	aggtgctacg	gtgcagctga	ctccatattt	tcctacttgt
ggcagcgact	gcatccgaca	taaaggaaca	gttggtgctc	gcccacaaac	aggcgtccct
ttccctctgg	ataacaacaa	aagcaagccg	ggaggctggc	tgcctctcct	cctgctgtct
ctgctggtgg	ccacatgggt	gctggtggca	gggatctatc	taatgtggag	gcacgaaagg
atcaagaaga	cttccttttc	taccaccaca	ctactgcccc	ccattaaggt	tcttgtgggt
tacctatctg	aaatatgttt	ccatcacaca	atttgttact	tactgaatt	tcttcaaaac
cattgcagaa	gtgaggtcat	ccttgaaaag	tggcagaaaa	agaaaatagc	agagatgggt

35

40

	ccagtgcagt	ggcttgccac	tcaaaagaag	gcagcagaca	aagtcgtctt	ccttctttcc
	aatgacgtca	acagtgtgtg	cgatgggtacc	tgtggcaaga	gcgagggcag	tcccagttag
	aactctcaag	actcttcccc	ttgcctttta	ccttttctgc	agtgatctaa	gaagccagat
	tcatctgcac	aaatacgtgg	tggctctactt	tagagagatt	gatacaaaaag	acgattacaa
5	tgctctcagt	gtctgcccc	agtaccacct	catgaaggat	gccactgctt	tctgtgcaga
	acttctccat	gtcaagtagc	aggtgtcagc	aggaaaaaga	tcacaagcct	gccacgatgg
	ctgctgctcc	ttgtagccca	cccatgagaa	gcaagagacc	ttaaaggctt	cctatccccac
	caattacagg	gaaaaaacgt	gtgatgatcc	tgaagcttac	tatgcagcct	acaaacagcc
	ttagtaatta	aaacatttta	taccaataaa	attttcaa	attgctaact	aatgtagcat
10	taactaacga	ttggaaacta	catttacaac	ttcaaagctg	ttttatacat	agaaatcaat
	tacagtttta	attgaaaact	ataaccattt	tgataatgca	acaataaagc	atcttcagcc
	aaaaaaaaa	aaaaaa				

AF208110

15	cggcgatgtc	gctcgtgctg	ataagcctgg	ccgcgctgtg	caggagcgcc	gtacccccgag
	agccgaccgt	tcaatgtggc	tctgaaactg	ggccatctcc	agagtggatg	ctacaacatg
	atctaattccc	cgggagacttg	agggacctcc	gagtagaacc	tgttacaact	agtgttgcaa
	caggggacta	ttcaattttg	atgaatgtaa	gctgggtact	ccgggcagat	gccagcatcc
	gcttgttgaa	ggccaccaag	atgtgtgtga	cgggcaaaaag	caacttccag	tcctacagct
20	gtgtgaggtg	caattacaca	gaggccttcc	agactcagac	cagaccctct	ggtggtaaat
	ggacattttc	ctatatcggc	ttccctgtag	agctgaacac	agtctatttc	attggggccc
	ataatattcc	taatgcaa	atgaatgaag	atggcccttc	catgtctgtg	aatttcacct
	caccaggctg	cctagaccac	ataatgaa	ataaaaaaaa	gtgtgtcaag	gccggaagcc
	tgtgggatcc	gaacatcact	gcttgtaaga	agaatgagga	gacagtagaa	gtgaacttca
25	caaccactcc	cctgggaaac	agatacatgg	ctcttatcca	acacagcact	atcatcgggt
	tttctcaggt	gtttgagcca	caccagaaga	aacaaacgcg	agcttcagtg	gtgattccag
	tgactgggga	tagtgaaggt	gctacggtgc	agctgactcc	atattttcct	acttgtggca
	gcgactgcat	ccgacataaa	ggaacagttg	tgctctgccc	acaaacaggc	gtccctttcc
	ctctggataa	caacaaaagc	aagccgggag	gctggctgcc	tctcctcctg	ctgtctctgc
30	tggtggccac	atgggtgctg	gtggcaggga	tctatcta	gtggaggcac	gaaaggatca
	agaagacttc	cttttctacc	accacactac	tgccccccat	taaggttctt	gtggtttacc
	catctgaaat	atgtttccat	cacacaattt	gttacttcac	tgaatttctt	caaaaccatt
	gcagaagtga	ggtcatcctt	gaaaagtggc	agaaaaagaa	aatagcagag	atgggtccag
	tgcagtggct	tgccactcaa	aagaaggcag	cagacaaagt	cgtcttcctt	ctttccaatg
35	acgtcaacag	tgtgtgcgat	ggtacctgtg	gcaagagcga	gggcagtccc	agtgagaact
	ctcaagacct	cttccccctt	gcctttaacc	ttttctgcag	tgatctaaga	agccagattc
	atctgcacaa	atacgtgggt	gtctacttta	gagagattga	tacaaaagac	gattacaatg
	ctctcagtgt	ctgccccaa	taccacttca	tgaaggatgc	cactgctttc	tgtgcagaac
	ttctccatgt	caagcagcag	gtgtcagcag	gaaaaagatc	acaagcctgc	cacgatggct
40	gctgctcctt	gtagcccacc	catgagaagc	aagagacctt	aaaggcttcc	tatcccacca
	attacaggga	aaaaacgtgt	gatgatcctg	aagcttacta	tgcagcctac	aaacagcctt
	agtaattaaa	acattttata	ccaataaaa	tttcaaata	tactaactaa	tgtagcatta
	actaacgatt	ggaaactaca	tttacaactt	caaagctgtt	ttatacatag	aatcaatta

cagctttaat tgaaaactgt aaccatttttg ataatgcaac aataaagcat cttccaaaaa
aaaaaaaaaa aaaaaaaaaa aaaaaaaa

AF208111

5	cggcgatgtc	gctcgtgctg	ataagcctgg	ccgcgctgtg	caggagcgcc	gtacccccgag
	agccgaccgt	tcaatgtggc	tctgaaactg	ggccatctcc	agagtggatg	ctacaacatg
	atctaattccc	cggagacttg	agggacctcc	gagtagaacc	tgttacaact	agtgttgcaa
	caggggacta	ttcaattttg	atgaatgtaa	gctgggtact	ccgggcagat	gccagcatcc
	gcttggttgaa	ggccaccaag	atltgtgtga	cgggcaaaag	caacttccag	tcctacagct
10	gtgtgaggtg	caattacaca	gaggccttcc	agactcagac	cagaccctct	gggtggtaaat
	ggacattttc	ctatatcggc	ttccctgtag	agctgaacac	agtctatttc	attggggccc
	ataatattcc	taatgcaaat	atgaatgaag	atggcccttc	catgtctgtg	aatttcacct
	caccaggctg	cctagaccac	ataatgaaat	ataaaaaaaa	gtgtgtcaag	gccggaagcc
	tgtgggatcc	gaacatcact	gcttgtaaga	agaatgagga	gacagtagaa	gtgaacttca
15	caaccactcc	cctgggaaac	agatacatgg	ctcttatcca	acacagcact	atcatcgggt
	tttctcaggt	gtttgagcca	caccagaaga	aacaaacgcy	agcttcagtg	gtgattccag
	tgactgggga	tagtgaaggt	gctacggtgc	aggtaaagtt	cagtgagctg	ctctggggag
	ggaagggaca	tagaagactg	ttccatcatt	cattgctttt	aaggatgagt	tctctcttgt
	caaatgcact	tctgccagca	gacaccagtt	aagtggcgtt	catgggggtt	ctttcgctgc
20	agcctccacc	gtgctgaggt	caggaggccg	acgtggcagt	tgtggtcctt	tttgcttgta
	ttaatggctg	ctgaccttcc	aaagcacttt	ttattttcat	tttctgtcac	agacactcag
	ggatagcagt	accattttac	ttccgcaagc	ctttaactgc	aagatgaagc	tgcaaagggg
	ttgaaatggg	aaggtttgag	ttccaggcag	cgtatgaact	ctggagaggg	gctgccagtc
	ctctctgggc	cgcagcggac	ccagctggaa	cacaggaagt	tggagcagta	gggtgctcctt
25	cacctctcag	tatgtctctt	tcaactctag	tttttgaagt	ggggacacag	gaagtccagt
	ggggacacag	ccactcccca	aagaataagg	aacttccatg	cttcattccc	tggcataaaa
	agtgntcaaa	cacaccagag	ggggcaggca	ccagccaggg	tatgatgggt	actacccttt
	tctggagaac	catagacttc	ccttactaca	gggacttgca	tgtcctaaag	cactggctga
	aggaagccaa	gaggatcact	gctgctcctt	ttttgtagag	gaaatgtttg	tgtacgtggg
30	aagatatgac	ctagcccttt	taggtaagcg	aactggatatg	ttagtaacgt	gtacaaagtt
	taggttcaga	ccccgggagt	cttgggcagt	tgggtctcgg	gtcactgggt	ttgacttttag
	ggctttgtta	cagatgtgtg	accaagggga	aatgtgtcat	gacaacacta	gaggtagggg
	cgaagccaga	aagaagggaa	gttttggtg	aagtaggagt	cttgggtgaga	ttttgctgtg
	atgcatgggt	tgaactttct	gagcctcttg	tttttcctca	gctgactcca	tattttccta
35	cttgtggcag	cgactgcata	cgacataaag	gaacagttgt	gctctgcccc	caaacaggcg
	tccctttccc	tctggataac	aacaaaagca	agccgggagg	ctggctgcct	ctcctcctgc
	tgtctctgct	gggtggccaca	tgggtgctgg	tggcagggat	ctatctaattg	tggaggcacg
	aaaggatcaa	gaagacttcc	ttttctacca	ccacactact	gccccccatt	aagggtcttg
	tgggtttacc	atctgaaata	tgtttccatc	acacaatttg	ttacttcaact	gaatttcttc
40	aaaaccattg	cagaagtgtg	gtcatccttg	aaaagtggca	gaaaaagaaa	atagcagaga
	tgggtccagt	gcagtggctt	gccactcaaa	agaaggcagc	agacaaagtc	gtcttccttc
	tttccaatga	cgtcaacagt	gtgtgcgatg	gtacctgtgg	caagagcgag	ggcagtccca
	gtgagaactc	tcaagacctc	ttcccccttg	cctttaacct	tttctgcagt	gatctaagaa

gccagattca tctgcacaaa tacgtggtgg tctacttttag agagattgat acaaaagacg
attacaatgc tctcagtgtc tgccccaagt accacttcat gaaggatgcc actgctttct
gtgcagaact tctccatgtc aagcagcagg tgtcagcagg aaaaagatca caagcctgcc
acgatggctg ctgctccttg tagccacccc atgagaagca agagacctta aaggcttccct
5 atcccaccaa ttacagggaa aaaacgtgtg atgacctga agcttactat gcagcctaca
aacagcctta gtaattaaaa cattttatata caataaaatt ttcaaataatt actaactaat
gtagcattaa ctaacgattg gaaactacat ttacaacttc aaagctgttt tatacataga
aatcaattac agctttaatt gaaaactgta accatthttga taatgcaaca ataaagcatc
ttccaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaa

10

AF250309

atgtcgctcg tgctgctaag cctggccgcg ctgtgcagga gcgccgtacc ccgagagccg
accgttcaat gtggctctga aactgggcca tctccagagt ggatgctaca acatgatcta
atcccgggag acttgaggga cctccgagta gaacctgtta caactagtgt tgcaacaggg
15 gactattcaa ttttgatgaa tgtaagctgg gtactccggg cagatgccag catccgcttg
ttgaaggcca ccaagatttg tgtgacgggc aaaagcaact tccagtccta cagctgtgtg
aggtgcaatt acacagaggc cttccagact cagaccagac cctctggtgg taaatggaca
ttttcctata tcggcttccc tgtagagctg aacacagtct atttcatttg ggcccataat
attcctaata caaatatgaa tgaagatggc cttccatgt ctgtgaattt cacctcacca
20 ggctgcctag accacataat gaaatataaa aaaaagtgtg tcaaggccgg aagcctgtgg
gatccgaaca tcaactgcttg taagaagaat gaggagacag tagaagtga cttcacaacc
actcccctgg gaaacagata catggctctt atccaacaca gcactatcat cgggttttct
caggtgtttg agccacacca gaagaaacaa acgcgagctt cagtggatgat tccagtgact
ggggatagtg aaggtgctac ggtgcagctg actccatatt ttctactttg tggcagcgac
25 tgcacccgac ataaaggaac agttgtgctc tgcccacaaa caggcgctcc tttccctctg
gataacaaca aaagcaagcc gggaggctgg ctgcctctcc tctgctgtc tctgctgggtg
gccacatggg tgctgggtggc agggatctat ctaatgtgga ggcacgaaag gatcaagaag
acttcctttt ctaccaccac actactgccc cccattaagg ttcttgtgggt ttacccatct
gaaatatgtt tccatcacac aatttgttac ttcactgaat ttcttcaaaa ccattgcaga
30 agtgaggtca tcttgaaaa gtggcagaaa aagaaaatag cagagatggg tccagtgcag
tggcttgcca ctcaaaagaa ggcagcagac aaagtcgtct tctttctttc caatgacgtc
aacagtgtgt gcgatggtac ctgtggcaag agcgagggca gtcccagtga gaactctcaa
gacctcttcc cccttgctt taaccttttc tgcagtgatc taagaagcca gattcatctg
cacaataacg tgggtggtcta ctttagagag attgatacaa aagacgatta caatgctctc
35 agtgtctgcc ccaagtacca cctcatgaag gatgccactg ctttctgtgc agaacttctc
catgtcaagc agcaggtgtc agcaggaaaa agatcacaa cctgccacga tggctgctgc
tccttgtagc ccacccatga gaagcaagag accttaaagg gttccttttc ccatcattta
caggggaaaa acgtgtgatg atc

40

AK095091

catattagag tctacagata tgccttttctt acagcaatcc tgcacccaca taaaagctac
atthttcaata caagattaaa aggtattctg caaatgtgc aaggthttca tgtctgctgg

tgtagctgta gtgatggctt catgaatfff tttctttttt gactatgggc cttacgctgg
attcatttat cttgaaatgg tgaacaatca cagctgcaga ccctcaattt atggtacata
tcaagcaatt tggctttttt tcttgtaatg aaaaaaaaaa gttttttttg ctttttttca
tgacactgct tcttgggagc actgccagca ttactagtgg cacttcgcat gggtcctaag
5 gtgttattga aggtttacga tattgcacta aacacgaaaa ataccagaga accactggag
atacttttta ctgtgatatg taatttactg gagacaggaa ctgctcgttt ggagatgggt
agcatcacag ggtgttttaa gtcgatactt gcaacccttg agctcaccac agtagcaaca
ggaggtggct aggaaattat tcacagcagg acagtacgca ctgcaattaa ttgtatgcag
ttatgattta ataccacatc tttatgctca cgtttctctc aactgtgaat ggtgccatgt
10 acagttggta tgtgtgtgtt taagttttga taaatfttta acttttaata gttaaaatag
ttaactattg gtatggtagg aaatgataaa gtagactagt atctgtatac attttctgca
tttatgacat acctttttct tcattttttt caatatttta attgaaaagt tcatccgagt
ttcatctaag ttttttcaaa gtgatacaaa tctccaaaaa attttccaat atatgtattg
aaaaaatcca ggtgtaagtg gctctgcgca gtccaaacct gtgttgttca aggtcaact
15 gtgtatgaat ccaagcgaaa gcttttctta acacctcata agaactattt tttaaaaaac
aggaactagc atagagtaac catcacaggt aaagtgtaat ttgttatcag ccatcttttg
cccatttcag tactggtaga aggtcfaatg gtaaaaaata aaacgggaca gtcagaagat
ctggaagtcg tgaccctgct ttcacctggc atgtgtaatc cagtcatgct cgtatcagtc
tctgtaggag cacttgaagg tattacataa atgctatcta actctgggaa acgccaacat
20 gtgattgcct ccagaggaat cttcttttaa aaaaaattca aaatgttatt tccttactag
gatgtcttta aagaattata acctttaccg tgctccaca ttagatagat ccctgccacc
agcacccatg tggccaccag cagagacagc aggaggagag gcagccagcc tcccggcttg
cttttgtctg gaaaaaacia agcttattca cctttggaaa acaaattcac acttatctct
taatttaaaa actaagactt ggtatacttt atagaggttt atttattttt tattattttt
25 tagttttgag acagagtctc gctttgttgc ctaggctgga gtgcagtggc gcaatctcgg
ttcactgcag cctccgtctc ccgggttcaa gcaatgctgc ctcagcctcc tgagtagctg
ggattacagg catgtgtcac cgcgccagc cactttgtag agatttagat ccctttaaaa
ccatcagtea gaagctcttt agatagtctg ccaatcatat ctttttccct agagtgtgca
ggtcttgcat tagattctca aaagggatat gggaccagc aagttaagaa cagtccataa
30 atctcttttg cttctttgtc ctgatatgca ccggcatttt cacagtagga actagggttt
ctgtccagtt tttttggttc ttttaaggaat taatgttatt ctgggtacaa ctgcttacat
acatagcaca tatagatgac atttttacag gccgtcttgt tagactgaca tacatggagg
atagtgccac ccgcctcaca agaacatcag gtaagctcag gcacagagtg cccaggaatc
tgtaaggctt cgcccacgca caagtcaggg ctgccagtca cctgggttgt cttcacttta
35 tttggctgcy tctaatagaca ccttccaact tttgaccca ccctggact gttgtgtaaa
cattgtattt ctccatctgt aatgaaaaag ctaacacatc tctaactcca gagacatttt
ccagaacatg ctgttctcag gcactagtga ggcggtacca ttattcctca tttgttatcc
aatgttggc catgtgacca caccaaaagc tcatcctggg ccactgagac tagtaattga
atcagaatat agtgaatat tcattctcat atatacccag ccatcttaca tctttggctt
40 ttttcagcag atccttgtgg cactcagaac atccattttg cactgtgtat ttttttccct
tctgtgtatc ctgctttgta aagagtcacg agtggtttta caaataaagc ctgttcttac
tcag

BM983744

TTTTTTTTTTTTTTTTCTGAGTAAGAACAGGCTTTATTTGTAAAACCACTCGTGACTCTT
TACAAAGCAGGATACACAGAAGGGAAAAAATACACAGTGCAAAATGGATGTTCTGAGTG
CCACAAGGATCTGCTGAAAAAAGCCAAAGATGTAAGATGGCTGGGTATATATGAGAATGA
ATATTTCACTATATTCTGATTCAATTACCAGTCTCAGTGGCCCAGGATGAGCTTTTGGTG
5 TGGTCACATGGCCAACATTTGGATAACAAATGAGGAATAATGGTACCGCCTCACTAGTGC
CTGAGAACAGCATGTTCTGGAAAATGTCTCTGGAGTTAGAGATGTGTTAGCTTTTTTCATT
ACAGATGGAGAAATACAATGTTTACACAACAGTCCAGGGGTGGGGTCAAAAGTTGGAAGG
TGTCATTAGACGCAGCCAAATAAAGTGAAGACAACCCAGGTGACTGGCAGCCCTGACTTG
TGCGTGGGCGAAGCCTTACAGATTCCTGGGCACTCTGTGCCTGAGCTTACCTGATGTTCT
10 TGTGAGGCGGGTGGCACTATCCTCCATGTATGTCAGTCTAACAAGACGGCCTGTAAAAAT
GTCATCTATATGTGCTATGTATGTAAGCAGTTGTACCCAGAATAACATTAATCCTCGTGC
CGAAT

CB305764

15 TTTTTTTTTTTTTTTTGGTTGGGCTGAAGATGCTTTATTATTGCATTATCAAAATGGTTATA
GTTTTCAATTAAAACTGTAATTGATTTCTATGTATAAAACAGCTTTGAAGTTGTAAATGT
AGTTTCCAATCGTTAGTTAATGCTACATTAGTTAGCAATATTTGAAAATTTTATTGGTAT
AAAATGTTTTAATTACTAAGGCTGTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACA
CGTTTTTTCCCTGTAATTGGTGGGATAGGAAGCCTTTAAGGTCTCTTGCTTCTCATGGGT
20 GGGCTACAAGGAGCAGCAGCCATCGTGGCAGGCTTGTGATCTTTTTCTGCTGACACCTG
CTGCTTGACATGGAGAAGTTCTGCACAGAAAGCAGTGGCATCCTTCATGAGGTGGTACTT
GGGGCAGACACTGAGAGCATTGTAATCGTCTTTTGTATCAATCTCTCTAAAGTAGACCAC
CACGTATTTGTGCAGATGAATCTGGCTTCTTAGATCACTGCAGAAAAGGTTAAAGGCAAG
GGGGAAGAGGTCTTGAGAGTTCTCACTGGGACTGCCCTCGCTCTTGCCACAGGTACCATC
25 GCACACACTGTTNACGTCATTGGAAAGAAGGAAGACGACTTTGTCTGCTGCCTTCTTTTG
AGTG

BM715988

TGGTTTTTGTTTTTTTTTTCATTTTCTGTTGGATTACAGAAAAAGAATGGGACCCATTTCAG
30 GTCTCGATTTCCAAAGGTAAAGATGGAAGGCTGGGCAGACTGGCTTTTGTACCTGACAT
GCCGTAGGGTGAGCTTAGAGGAAGAAAGAAAACAATTTTTATTTGGCCAAAACAGAACAA
ATGCTGAAAAGGAAATCTTGTTTTTTTCTAAAGCCAAATAGAAATGATTTGGGTATAAT
TTAAGAGTCCTTGTGTTGTACAGATATGGTGACTGATGTAGTTATTAATACTACCAACTT
AGTCATCAAGCCTCAATTTTCTTTACCTGAAGGATTAAGTGAAAGCTTTTGGAGTTTCAT
35 GATGTTTCAGTATGATCAGTTAACCTTAACCTCTGAGCATCCTGAAGCAAAATCTAAATAA
TGCAGCTATTACCACTGGTGGTCCAGGCTCTGGTGAAGCCCTCTGAGCCCAGGAGGAAGA
GAAAGCATTGTCCAGAGGTAGGAACACAGTCTGGGAGCCCAGAGCTCTGGGAGGAGTGGG
AAAATGCTGCTTCCCTGCTGCTTGCTTCTAGGCACCTGCTTCCGCCATCTCACTTACCATG
GCTAGAGATGGGGGTGAGACTGGGGAAGGACACAAGCAGGGAACAGATAAGGGATGGAAA
40 TCAGAAGGGAATATAGAAAGAACTCTGGATGTGGAGACATGCCGGTACCTGAGCATTTTG
TATCAATGGGAGTACCTCT

BM670929

TTTTTTTTTTTTTTTTTTGGCTGAAGATGCTTTATTGTTGCATTATCAAAATGGTTACAG
TTTTCAATTAAAGCTGTAATTGATTTCTATGTATAAAACAGCTTTGAAGTTGTAAATGTA
GTTTCCAATCGTTAGTTAATGCTACATTAGTTAGCAATATTTGAAAATTTTATTGGTATA
5 AAATGTTTTAATTACTAAGGCTGTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACAC
GTTTTTTTCCCTGTAATTGGTGGGATAGGAAGCCTTTAAGGTCTCTTGCTTCTCATGGGT
GGGCTACAAGGAGCAGCAGCCATCGTGGCAGGCTTGTGATCTTTTTCCCTGCTGACACCTG
CTGCTTGACATGGAGAAGTTCTGCACAGAAAGCAGTGGCATCCTTCATGAGGTGGTACTT
GGGGCAGACACTGAGAGCATTGTAATCGTCTTTTGTATCAATCTCTCTAAAGTAGACCAC
10 CACGTATTTGTGCAGATGAATCTGGCTTCTTAGATCACTGCAGAAAAGGTTAAAGGCAAG
GGGAAGAGGTCTTGAGAGTTCTCACTGGGACTTGCCCTCGCTCTTGCCACAGGTACCATC
GCACACACTGTTGACGTCATTGGAAAGAAAGAAGACGACTTTGTCTGCTGCCTTCTT

BI792416

15 GCTGAAGATGCTTTATTGTTGCATTATCAAAATGGTTACAGTTTCAATTAAAGCTGTAA
TTGATTTCTATGTATAAAACAGCTTTGAAGTTGTAAATGTAGT

BI715216

20 CACGCGTCCGATTTTATACCAATAAAATTTTCAAATATTGCTAACTAATGTAGCATTAAAC
TAACGATTGGAACTACATTTACAACCTCAAAGCTGTTTTATACATAGAAATCAATTACA
GCTTTAATTGAAAAGTGTAACTTTTGATAATGCAACAATAAAGCATCTTCAGCCAAAA
AAAAAA

N56060

25 AGAAAAAGAAAATAGCAGAGATGGGTCCAGTGCAGTGGCTTGCATAAAAAAGAAGGCAGC
AGACAAAGTCGTCTTCCTTCTTTCCAATGACGTCAACAGTGTGTGCGATGGTACCTGTGG
CAAGAGCGAGGGCAGTCCCAGTGAGAACTCTCAAGACCTCTTCCCCCTTGCCTTTAACC
TTTTCTGCAGTGTCTAAGAAGCCAGATTCATCTGCACAAATACGTGGTGGTCTACTTTA
GAGAGATTGATACAAAAGACGATTACAATGCTCTCAGTGTCTGCCCCAAGTACCACCTCA
30 TGAAGGATGCCACTGCTTTCTGTGCAGAACTTCTCCATGTCAAGCAGCAGGTTTCAGCAG
G

CB241389

35 TTTTTTTTTTTTTTTGTTTGGCTGAAGATGCTTTATTGTTGCATTATCAAAATGGTTACAG
TTTTCAATTAAAGCTGTAATTGATTTCTATGTATAAAACAGCTTTGAAGTTGTAAATGTA
GTTTCCAATCGTTAGTTAATGCTACATTAGTTAGCAATATTTGAAAATTTTATTGGTATA
AAATGTTTTAATTACTAAGGCTGTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACAC
GTTTTTTCCCTGTAATTGGTGGGATAGGAAGCCTTTAAGGTCTCTTGCTTCTCATGGGTG
GGCTACAAGGAGCAGCAGCCATCGTGGCAGGCTTGTGATCTTTTTCCCTGCTGACACCTGC
40 TGCTTGACATGGAGAAGTTCTGCACAGAAAGCAGTGGCATCCTTCATGAGGTGGTACTTG

GGGCAGACACTGAGAGCATTGTAATCGTCTTTTGTATCAATCTCTCTAAAGTAGACCACC
ACGTATTTGTGCAGATGAATCTGGCTTCTTAGATCACTGCAGAAAAGGTTAAAGGCAAGG
GGGAAGAGGTCTTGAGAGTTCTCACTGGGACTGCCCTCGCTCTTGCCACAGGTACCATCG
CACACACTGTTGACGTCAATTGGAAAGAAGGAAGACGACTTTGTCTGCTGCCTTCTTTTGA
5 GTGGCAAGCCACTGCACTGGACCCATCTCTGCTATTTTCTTTTCTNGCACTTTTCAAGG
ATGACTCACTTCTGCAATGGTTTTTGAGAATTCAGTGAAGTACAAATGTGTGATGGAACA
TAT

AV660618

10 CGCTCGTGCTGCTAAGCCTGGCCGCGCTGTGCAGGAGCGCCGTACCCCGAGAGCCGACCG
TTCAATGTGGCTCTGAAACTGGGCCATCTCCAGAGTGGATGCTACAACATGATCTAATCC
CCGGAGACTTGAGGGACCTCCGAGTAGAACCTGTTACAAC TAGTGTGCAACAGGGGACT
ATTCAATTTTGATGAATGTAAGCTGGGTACTCCGGGCAGATGCCACACCAGAAGAAACAA
ACGCGAGCTTCAGTGGTGATTCCAGTGACTGGGGATAGTGAAGGTGCTACGGTGCACTG
15 ACTCCATATTTTCCCTACTTGTGGCAGCGACTGCATCCGACATAAAGGAACAGTTGTGCTC
TGCCACAAACAGGCGTCCCTTTCCCTCTGGATAACAAC

BX088671

20 GCTGAGTGTGATGGTGTAAGCCTGTGGTCCCAGCTACTAGGGAGGCTGAGATGGGATTAC
AGGTGTGAGCCACGGCGCCTGGCCTAAAAGCATCTTTTTCTTTAACGCAGAGGTTATGTT
GTATTATTAGCATAAATGTTTTTTCTGGGAATGCTTATTTACACAGCACAATACTGAA
TCTTCTCTGGAATGTGGATCGATTTTCAGATGGATGACTATTAAAATGTGTATATTTGCAG
ATTATCCTTAAAGGGGCCACCTCATGCCTTCTAATTTATGTCTTACGGATAAAAAATCAAA
ATGAAGCATAAAGTAAAACTGTGTCCAGCTTTACAAGTGACGCTTAGTAATGGCTGAG
25 GCAATATGTTTAAATGTAGCAAATTTTACTTATTTGTGATGATCAGTTTTTCACAGTGCTTG
TAAGTGCTGGTAATAGAAGATGGACATGGTTTAGGTCAAACTTGGAACAGAAACCAACT
TCCTTTGAAACAGCTCTACCAGNTATAAGAGCAATATG

CB154426

30 CTGTTGACGTCAATTGGAAAGAAGGAAGACGACTTTGTCTGCTGCCTTCTTTTGAGTGGCA
AGCCACTGCACTGGACCCATCTCTGCTATTTTCTTTTCTGCCACTTTTCAAGGATGACC
TCACTTCTGCAATGGTTTTTGAAGAAATTCAGTGAAGTAACAAATTGTGTGATGGAAACAT
ATTTTCAGATGGGTAAACCACAAGAACCTTAATGGGGGGCAGTAGTGTGGTGGTAGAAAAG
GAAGTCTTCTTGATCCTTTCTGTGAGAGGAGAAAAGCATTGTGTTATCTGTGAACAGCAAA
35 CAGCAGGCTTTCACTCTGTAAACCATCCCTGACAAATGATCCCTTGCTAGAGAATGTCAG
CTGAGCACCAAGGGCCTTGTTAGTGACAGCAAGGAAAAACATCCTGATGTTCTTTTGAA
CACATCACCTGAAACACACTGATGCTTAAACCTTAACTTTTTTTTTTTTGGAGACACAGT
CTCACTCTGT

40 **CA434589**

TTTTTTTTTTTTTTTTTTTTCTGAGTAAGAACAGGCTTTATTTGTAAAACCACTCGTGACTC
TTTACAAAGCAGGATACACAGAAGGGAAAAAATACACAGTGCAAAATGGATGTTCTGAG
TGCCACAAGGATCTGCTGAAAAAAGCCAAAGATGTAAGATGGCTGGGTATATATGAGAAT
GAATATTTCACTATATTCTGATTCAATTACCAGTCTCAGTGGCCCAGGATGAGCTTTTGG
5 TGTGGTCACATGGCCAACATTTGGATAACAAATGAGGAATAATGGTACCGCCTCACTAGT
GCCTGAGAACAGCATGTTCTGGAAAATGTCTCTGGAGTTAGAGATGTGTTAGCTTTTTCA
TTACAGATGGAGAAATACAATGTTTACACAACAGTCCAGGGGTGGGGTCAAAAGTTGGAA
G

10 CA412162

TTTTTTTTTTTTTTTTTTTTGGCTGAAGATGCTTTATTGTTGCATTATCAAAATGGTTATAG
TTTTCAATTAAACTGTAAATTGATTTCTATGTATAAAACAGCTTTGAAGTTGTAAATGTA
GTTTCCAATCGTTAGTTAATGCTACATTAGTTAGCAATATTTGAAAATTTTATTGGTATA
AAATGTTTAAATTACTAAGGCTGTTTGTAGGCTGCATAGTAAGCTTCAGGATCATCACAC
15 GTTTTTTCCCTGTAATTGGTGGGATAGGAAGCCTTTAAGGTCTCTTGCTTCTCATGGGTG
GGCTACAAGGAGCAGCAGCCATCGTGGCAGGCTTGTGATCTTTTTCCTGCTGACACCTGC
TGCTTGACATGGAGAAGTTCTGCACAGAAAGCAGTGGCATCCTTCATGAGGTGGTACGTG
GGGCAGACACTGAGAGCATTTGTAATCGTCTTTTGTATCAATCTCTCTAAAGTAGACCACC
ACGTATTTGTGCAGATGAATCTGGCTTCTTAGATCACTGCAGAAAAGGTAAAGGCAAGG
20 GGGAAGA

CA314073

TTTTTTTTTTTTTTTTTTTTGAAAGGGTCAGGACTTCCAGATCTTCTGACTGTCCCGTTTTT
ATTTTTACCATTGAGCCTTCTACCAGTACTGAAATGGGCAAAAGATGGCTGATAACAAAT
25 TACACTTTACCTGTGATGGTTACTCTATGCTAGTTCCTGTTTTTTAAAAAATAGTTCTTA
TGAGGTGTTAAGAAAAGCTTTCGCTTGGATTTCATACACAGTTGACCCTTGAACAACACAG
GTTTGGACTGCGCAGAGCCACTTACACCTGGATTTTTTTCAATACATATATTGGAAAATTT
TTTGGAGATTTGTATCACTTTGAAAAAAGCTTAGATGAAACTCGGATGAACTTTTCAATTA
AAATATTGAAAAAATGAAGAAAAAGGTATGTCATAAATGCAGAAAATGTATACAGATAC
30 TAGTCTACTTTATCATTTTCTACCATAACCAATAGTTAACTATTTTAACTATTAAAAGTTA
AAAATTTATCAAACTTAAACACACACATACCAACTGTACATGGCACCATTACAGTTGA
GAGAAACGTGAGCATAAAGATGTGGTATTAAATCATAACTGCATACAATTAATTGCAGTG
CGTACTGTCTGTGCTGTGAATATTTCTAGCCCTCGTGCCGAATC

35 BF921554

GTGGGTGACCGTGGCTTGCCACTCAAAAGAAGGCAGCAGACAAAGTCGTCTTCCTTCTTT
CCAATGACGTCAACAGTGTGTGCGATGGTACCTGTGGCAAGAGCGAGGGCAGTCCAGTG
AGAACTCTCAAGACCTCTTCCCCCTTGCCCTTAACTTTTCTGCAGTGATCTAAGAAGCC
AGATTCATCTGCACAAATACGTGGTGGTCTACTTTAGAGAGATTGATACAAAGACGATT
40 ACAATGCTCTCAGTGTCTGCCCCAAGTACCACCTCATGAAGGATGCCACTGCTTTCTGTG
CATAACTTCTCCATGTCAAGCAGCAGGTGTGAGCAGGAAAAAGATCACAAGCCTGCCACG

ATGGCTGCTGCTCCTTGTAGCCCACCCATGAGAAGCAAGAGACCTTAAAGGCTTCCTATC
CCACCAATTACAGGGAAAAAACGTGTGATGATCCTGAAGCCACGGTCAA

BF920093

5 TAGAGGATCCCGGTGACGGTGGTTCAGTGATCATCACACTTTTCCCTGTAATAGGTGG
GATAGGAAGCCTTTAAGGTCTCTTGCTTCTCATGGGTGGGCTACAAGGAGCAGCAGCCAT
CGTGGCAGGCTTGTGATCTTTTTCCTGCTGACACCTGCTGCTTGACATGGAGAAGTTATG
CACAGAAAGCAGTGGCATCCTTCATGAGGTGGTACTTGGGGCAGACACTGAGAGCATTGT
AATCGTCTTTTGTATCAATCTCTCTAAAGTAGACCACCACGTATTTGTGCAGATGAATCT
10 GGCTTCTTAGATCACTGCAGAAAAGGTTAAAGGCAAGGGGGAAGAGGTCTTGAGAGTTCT
CACTGGGACTGCCCTCGCTCTTGCCACAGGTACCATCGCACACACTGTTGACGTCATTGG
AAAGAAGGAAGACGACTTTGTCTGCTGCCTTCTTTTGAGTGGCAAGCCACGGTCAACCCA
CAAGCCACGGTCAACCCAC

15 AV685699

TCTACGTGGTAAGATATGACCTAGCCCTTTTAGGTAAGCGAACTGGTATGTTAGTAACGT
GTACAAAGTTTtaggtTCAGACCCCGGGAGTCTTGGGCATGTGGGTCTCGGGTCACTGGTT
TTGACTTTAGGGCTTTGTTACAGATGTGTGACCAAGGGGAAAATGTGCATGACAACACTA
GAGGTAGGGGCGAAGCCAGAAAGAAGGGAAGTTTGGCTGAAGTAGGAGTCTTGCGACTG
20 CATCCGACATAAAGGAACAGTTGTGCTCTGCCCCACAAACAGGCGTCCCTTTCCCTCTGGA
TAACAACAAAAGCAAGCCGGGAGGCTGGCTGCCTCTCCTCCTGCTGTCTCTGCTGGTGGC
CACATGGGTGCTGGTGGCAGGGATCTATCTAATGTGGAGGCACGAAAGGATCAAGAAGAC
TTCTTTTTCTACCACCACACTACTGCCCCCATTAAGGTTCTTGTTGGTTTACCCATCTGA
AATATGTTTCCATCACACAATTTGTTACTTCACTGAATTTCTTCAAAACCATTGCAGAAG
25 TGAGGTCATCCTTGAAAGTGGCAGAGTAGCAGAGATGGGTCCAGTGCAGTGGCTTGCCAC
TCGTGCGATGGTCTT

AV650175

GGCACGAGCACTGGCTGAAGGAAGCCAAGAGGATCACTGCTGCTCCTTTNTTCTAGAGGA
30 AATGTTTGTCTACGTGGTAAGATATGACCTAGCCCTTTTAGGTAAGCGAACTGGTATGTT
AGTAACGTGTACAAAGTTTtaggtTCAGACCCCGGGAGTCTTGGGCATGTGGGTCTCGGGT
CACTGGTTTTGACTTTAGGGCTNTGTTACAGATGTGTGACCAAGGGGAAAATGTGCATGA
CAACACTAGAGCTGACTCCATATTTTCCCTACTTGTGGCAGCGACTGCATCCGACATAAAG
GAACAGTTGTGCTCTGCCCCACANACAGGCGTCCCTTTCCCTCTGGATAACAACATAAGCA
35 AGCCGGGAGGCTGGCTGCCTCTCCTCCTGCTGTCTCTGCTGGTGGCACATGGGTGCTGGT
GGAGGGATCTATCTAATGTGGAGGCACGGATCAAGAAGACTTNCTTNTCTACCACCACAC
TACTGGCCCCAATAAGGGTCTNGTGGNTACCCCATCTGAATATGTTTATACACAATTTGT
ACTCACTGAATTCTCAAAACATTGAGAGTGAGGCATCCTGAAAGTGCGAAAAGANATGCN
AATGGTCAGTGCATGCTGCACTAGCAGCATGGACTT

40

BX483104

GATCCCGCGCAGTGGCCCGGCGATGTCGCTCGTGCTGCTAAGCCTGGCCGCGCTGTGCAG
GAGCGCCGTACCCCGAGAGCCGACCGTTCAATGTGGCTCTGAAACTGGGCCATCTCCAGA
GTGGATGCTACAACATGATCTAATCCCGGAGACTTGAGGGACCTCCGAGTAGAACCTGT
TACAAC TAGTGTTGCAACAGGGGACTATTCAATTTTGATGAATGTAAGCTGGGTACTCCG
5 GGCAGATGCCAGCATCCGCTTGTTGAAGGCCACCAAGATTTGTGTGACGGGCAAAAGCAA
CTTCCAGTCCTACAGCTGTGTGAGGTGCAATTACACAGAGGCCTTCCAGACTCAGACCAG
ACCTCTGGTGGTAAATGGACATTTTCCTACATCGGCTTCCCTGTAGAGCTGAACACAGT
CTATTTCAATTGGGGCCCATATATTCCTAATGCAAATATGAATGAAGATGGCCCTTCCAT
GTCTGTGAATTTACCTCACCAGGCTGCCTAGACCACATAATGAAATATAAAAAAAGTG
10 TGTCAAGGCCGGAAGCCTGTGGGATCCGAACATCACTGCTTGTAAGAAGAATGAGGAGAC
AGTAGAAGTGAACCTTACAACCACTCCCTGGGAAACAGATACATGGCTCTTATCCAACA
CAGCACTATCATTCGG

CD675121

15 GTCTTGCATTAGATTCTCAAAAGGGATATGGGACCCAGGAAGTTAAGAACAGTCCTAAAA
TCTCTTTGGCTTCTTTGTCCTGATATGCACCGGCATTTTCACAGTAGGAAGTAGGGTTTC
TGTCCAGTTTTTTTTGGTTCTTTAAGGAATTAATGTTATTCTGGGTACAAGTACTGCTTACATA
CATAGCACATATAGATGACATTTTACAGGCCGTCTTGTTAGACTGACATACATGGAGGA
TAGTGCCACCCGCCTCACAAGAACATCAGGTAAGCTCAGGCACAGAGTGCCCAGGAATCT
20 GTAAGGCTTCGCCCACGCACAAGTCAGGGCTGCCAGTCACCTGGGTGTCTTCACTTTAT
TTGGCTGCGTCTAATGACACCTTCCAACCTTTTGACCCACCCCTGGACTGTTGTGTAAAC
ATTGTATTTCTCCATCTGTAATGAAAAAGCTAACACATCTCTAACTCCAGAGACATTTTC
CAGAACATGCTGTTCTCAGGCACTAGTGAGGCGGTACCATTATTCCTCATTTGTTATCCA
AATGTTGGCCATGTGACCACACCAAAAGCTCATCCTGGGCCACTGAGACTGGTAATTGAA
25 TCAGAATATAGTGAAATATTCATTCTCATATATACCCAGCCATCTTACATCTTTGGCTTT
TTTCAGCAGATCCTTGTGGCACTCAGAACATCCATTTTGCACTGTGTATTTTTTCCCTT
CT

BE081436

30 TGTGTAAGTCTCAAGACCTCTTCCCCCTTGCCTTTAACCTTTTCTGCAGTGATCTAAGAA
GCCAGATTCATCTGCACAAATACGTGGTGGTCTACTTTAGAGAGATTGATACAAAAGACG
ATTACAATGCTCTCAGTGTCTGCCCCAAGTACCACCTCATGGAGGATGCCACTGCTTTCT
GTGCAGAACTTCTCCATGTCAAGTAGCAGGTGTCAGCAGGAAAAAGATCACAAGCCTGCC
ACGATGGCTGCTGCTCCTTGTAGCCCACCCATGAGAAGCAAGAGACCTTAAAGGCTTCCT
35 ATCCCACCAATTACAGGGAAAAAACGTGTGATGAT

AW970151

CTGAAATATGTTTCCATCACACAATTTGTTACTTCACTGAATTTCTTCAAAACCATTTGCA
GAAGTGAGGTCATCCTTGAAAAGTGGCAGAAAAAGAAAATAGCAGAGATGGGTCCAGTGC
40 AGTGGCTTGCCACTCAAAAGAAGGCAGCAGACAAAGTCGTCTTCCTTCTTTCCAATGACG
TCAACAGTGTGTGCGATGGTACCTGTGGCAAGAGCGAGGGCAGTCCCAGTGAGAACTCTC

AAGACCTCTTCCCCCTTGCCTTTAACCTTTTCTGCAGTGATCTAAGAAGCCAGATTTCATC
TGCACAAATACGTGGTGGTCTACTTTAGAGAGATTGATACAAAAGACGATTACAATGCTC
TCAGTGTCTGCCCCAAGTACCACCTCATGAAGGATGCCACTGCTTTCTGTGCAGAACTTC
TCCATGTCAAGTAGCAGGTGTCAGCAGGAAAAAGATCACAAGCCTGCCACGATGGCTGCT
5 GCTCCTTGTAGCCCAACCATGAGAAGCAAGAGACCTTAAAGGCTTCCTATCCCACCAATT
ACAGGGAAAAAACGTGTGATGATCCCTGAAGCTTACTATGCAGCCTACANACAGCCTTA
GTAATAAAACATTTTATCCAATAAAATTTCAAATTTTGCTTAACTATGTGCATAAACTAC
GATTGAAAACCTTTTACACT

10 AW837146

CATTGTGGTTGCAGCTGCATAGTAAGCTTCAGGATCATCACACGTTTTTTCCCTGTAATT
GGTGGGATAGGAAGCCTTTAAGGTCTCTTGCTTCTCATGGGTGGGCTACAAGGAGCAGCA
GCCATCGTGGCAGGCTTGTGATCTTTTTCCCTGCTGACACCTGCTGCTTGACATGGAGAAG
TTCTGCACAGAAAGCAGTGGCATCCTTCATGAGGTGGTACTTGGGGCAGACACTGAGAGC
15 ATTGTAATCGTCTTTTGTATCAATCTCCCTAAAGTAGACCACCACGTATTTGTGCAGATG
AATCTGGCTTCTTAGATCACTGCAGAAAAGGTTAAAGGCAAGGGGAAGAGGTCTTGAGA
GTTCTCACTGGGACTGCCCTCGCTCTTGCCACAGGTACCATCGCACACACTGTTGACGTC
ATTGGAAGAAGGAAGACGACTTTGTCTGCTGCCTTCTTTTGAGTGGCAAGCCACTGCAC
TGGACCCATCT

20

AW368264

GTGAATAAGCTTTGTTTTTTCCAGACAAAAGCAAGCCAGGAGGCTGGCTGCCTCTCCTCC
TGCTGTCTCTGCTGGTGGCCACATGGTTGCTGGTGGCAGGGATCTATCTAATGTGGAGGC
ACGGTAAGGGTTATAATTCTTTAAAGTCATCCTAGTAAGGAAATAACATTTGGAATTTTT
25 TTTTAAAGAAGATTCCTCTGGAGGCAATCACCTGTTGGCGTTTCCCAGAGTTAGATAGCA
TTTATGTAATACCTTCAAGTGCTCCTACAGAGACTGATACGAGCATGACTGGATTACACA
TGCCAGGTGAAAGCAGGGCCAGGACTTCCAGATCTTCTGACTGTCCCGTTTTTTATTTTTA
CCATTGAGCCTTCTACCAGAACTGAAATGGGCAAAAGATGGCTGATAACAAATTACACTT
TACCTGTGATGGTTACTCTATGCTAGTTCCTGTTTTTAAAAAATAGTTCTTATGAGGTG
30 TCAAGAAAAGCTTTCGCTTGGAATTCATACACAGTTGACCCTTGAACAACACAG

D25960

GATCCTGAAGCTTACTATGCAGCCTACAAACAGCCTTAGTAATTAAAACATTTTATACCA
ATAAAATTTTCAAATATTGCTAACTAATGTAGCATTAACCTAACGATTGGAAACTACATNN
35 ACAACTTCAAAGCTGTTTTATACATAGAAATCAATTACAGCTTTAATTGAAAACATAAC
CATTTTGATAATGCAACANTAAAGCATCTTCAGCCAAA

AV709899

GCAACTTCCAGTCCTACAGCTGTGTGAGGTGCAATTACACAGAGGCCTTCCAGACTCAGA
40 CCAGACCCTCTGGTGGTAAATGGACATTTTCCCTATATCGGCTTCCCTGTAGAGCTGAACA

CAGTCTATTTTCATTGGGGCCCATAATATTCCTAATGCAAATATGAATGAAGATGGCCCTT
CCATGTCTGTGAATTTACCTCACCAGGCTGCCTAGACCACATAATGAAATATAAAAAA
AGTGTGTCAAGGCCGGAAGCCTGTGGGATCCGAACATCACTGCTTGTAAGAAGAATGAGG
AGACAGTAGAAGTGAACCTTCACAACCACTCCCCTGGGAAACAGATACATGGCTCTTATCC
5 AACACAGCACTATCATCGGGTTTTCTCAGGTGTTTGAGCCACACCAGAAGAAACAAACGC
GAGCTTCAGTGGTGATTCCAGTGACTGGGGATAGTGAAGGTGCTACGGTGCAGCTGACTC
CATATTTTCTACTTGTGGCAGCGACTGCATCCGACATAAAGGAACAGTTGTGCTCTGCC
CACAAACAGGCGTNCCTTTTCTCTGGATAACAACAAAAGCAAGCCGGGAGGCTTGGCTG
CTCTCCTTCTGCTGGCCTTTGCTGTGGCCACATTGGTGCTGGTGGCAGGGATCTATCTAA
10 TGTGGATGCACGTCTCGTGGTTTACCCATCTGAAATATGTTCN

BX431018

ATTTTTCTCTTGTGGCAGCGACTGGCATCCGACATAAAGGAACAGTTGTGCTCTGCCCCA
CAAACAGGCGTCCCTTTCCCTCTGGATAACAACAAAAGCAAGCCGGGAGGCTGGCTGCCT
15 CTCTCCTGCTGTCTCTGCTGGTGGCCACATGGGTGCTGGTGGCAGGGATCTATCTAATG
TGGAGGCACGAAAGGATCAAGAAGACTTCCTTTTCTACCACCACACTACTGCCCCCATT
AAGGTTCTTGTGGTTTACCCATCTGAAATATGTTTCCATCACACAATTTGTTACTTCACT
GAATTTCTTCAAACCATTTGCAGAAGTGAGGTCATCCTTGAAAAGTGGCAGAAAAAGAAA
ATAGCAGAGATGGGTCCAGTGCAGTGGCTTGCCACTCAAAGAAGGCAGCAGACAAAGTC
20 GTCTTCCTTCTTTCCAATGACGTCAACAGTGTGTGCGATGGTACCTGTGGCAAGAGCGAG
GGCAGTCCCAGTGAGA ACTCTCAAGACCTCTTCCCCCTTGCCTTTAACCTTTTCTGCAGT
GATCTAAGAAGCCAGATTTCATCTGCACAAATACGTGGTGGTCTACTTTAGAGAGATTGAT
ACAAAAGACGATTACAATGCTCTCAGTGTCTGCCCCAAGTACCACCTCATGAAGGATGCC
ACTGCTTTTCTGTGCAGAACTTCTCCATGTCAAGCAGCAGGTGTCAGCAGGAAAAAGATCA
25 CAAGCCTGCCACGATGGCTGCTGCTCCTTGTAGCCACCCATGAGAAGCAAGAGACCTTA
AGGCTTCTATCCCACCANTACAGGNAAAAACGTGTGATGATCCTGAAGCTTACTATGCAG
CCTACAACAGGCTTAGTATTAAACATTTTATACCCATAAATTTTCAAATTGCT

AL535617

TAGGTGACACTATAGAACAAGTTTGTACAAAAAAGCAGGCTGGTACCGGTCCGGAATTCC
CGGGATAGTGMCCGGCGAKGTCGCTCGTGCTGCTAAGCCTGGCCGCGCTGTGCAGGAGC
GCCGTACCCCGAGAGCCGACCGTTCAATGTGGCTCTGAACTGGGCCATCTCCARAGTGG
ATGSKACAACATGATCTAATCCCGGGAGACTTGAGGGACCTCCGAGTAGAACCTGTTACA
ACTAGTGTGCAACAGGGGACTATTCAATTTTGATGAATGTAAGCTGGGTACTCCGGGSA
35 GATGCCAGCATCCGCTTGTTGAAGGCCACCAAGATTTGTGTGAMGGGCAAAGCAACWTC
CAGTCCTACAGCWGTGTGAGGTAGCAATTACACAGAGAGCACATATCCAGACTCTAGACC
AGACCCTCTGGWGGTAAATGGACATTTTCTATATCGGCTTCCCTGTAGAGCTGAACACA
GTCTATATTCAATTGGGGCCCAWAATAWWCCTAATGCAAATATGAATGAAGATGGCCCTTC
CATGTCTGTGAATTTACCTCACCAGGCTGCCTAGACCACATAATGAAATAWAAAAAAA
40 GTGTGTCAAGGCCGGAAGCCTGTGGGATCCGAACATCACTGCTTGTAAGAAGAATGARGA
GACAGTAGAAGTGAACCTTCACAACCACTCCCCTGGGAAACAGATAMATKGCTCTTATCCA
ACACARMACTATCATCGGGTTTTCTCAGGTGTTTGAGCCACACCAGAAGAAACAAACGCG

AGCTTCAGTGGTGATTCCAGTGACTGGGGATAGTGAAGGTGCTACGGTGCAGCTGACTCC
ATATTTTCTACTTGTGGCAGCGWCTGCATCCGACATAAAGGAACAGTTGTGCTCTGCCC
ACAAACAGGCGTCCCTTTYCCTCTGGATAACAACAAAAGCAACYGGGAGSTGGYTGCT

5 AL525465

WAATWAKADDRATANHTGAAACTATAACCATTTNTGATAATNGNAANAATAAAGCATCT
TCAGCCAAACATCTAGTCTTCCATAGACCATGCATTGCAGTGTACCCAGAWCTGTTTAGC
TAATATTCTATGTTTAATTAATGAATACTAACTCTAAGAACCCCTCACTGATTCACCTCAA
TAGCATCTTAAGTGAAAAACCTTCTATTACATGCAAAAAATCATTGTTTTTAAGATAACA
10 AAAGTAGGGAATAAACAAGCTGAACCCACTTTTACTGGACCAAATGATCTATTATATGTG
TAACCACTTGTATGATTTGGTATTTGCATAAGACCTTCCCTCTACAACTAGATTCATAT
CTTGATTCTTGTACAGGTGCCTTTTAACATGAACAACAAAATACCCACAACTTGTCTAC
TTTTGCCTAAAGTTACCTATTAGAGGTCACTGTSAGAGTKCTCAGTTTCTTAGTTACTAT
TTAASTTTTSATGTTCAAAATGAAAATAATTCTKAAGTKGAAAGSGCTCTTGAAGTAACC
15 TTTTTATAAATGAGTTATTATAATGGTTTACTTAAATAAAAVAGAGGGGKTTTTGCGGTG
GCTCATGCCTCCAATCCCAGCACTTTGGCAAGGCCAAGGCAAAAVGATCGCTCAAGACCA
GGCTACGTCACAAAGCGAGACCTCCATCTCTACAAAAGATTTAAAAAATTAGCTGAGTGT
GATGGTGTGAGCCTGTGGTCCCAGCTACTAGGGAGGCTGAGATGGGAGGATCACTTGAGC
CCTGGAGGTCAAGGGTGCAGTAAACGGTGATTGTGCCACTGCACTCCATCCTGGGTGAGA
20 GCAGACCCTGTCTAAAACAAACAAACGAAAAAACCCCCACAGAATGACAGAACATAAAAG
ATGCACATTTTGTCTTCCAACCTTTTACTCTTCTAAAAGCATCTTTTTTAAATTTTTTAA
ATTTTTTTTTTTTTTGAGACAGAGTTTCACTCTGTACACAGGCTGGAGTGMGTGGCGTGA
CTCGGCTCACTAMAACCTCTGCTCCGGGGTYACSCATCTCCTGCWCAGCTCCTGAGAAGC
KGGAYAMAGGMCCACACAAACCAGTAAYTTTATWTTTTGAAAAAGGGTTYACCTGTASMA
25 GRAGGCTGAATCCGACMAARTMACCMCCACYYCAAADGAGGAWAAGKGKRSMSGSCBGGC
A

BX453536

TTATGGGGGGCAGTAGTGTGGTGGTAGAAAAGGAAGTCTTCTTGATCCTTTCGTGCCTCC
30 CATTAGATAGATCCCTGCCACCAGCACCCATGTGGCCACCAGCAGAGACAGCAGGAGGAG
AGGCAGCCAGCCTCCCGGCTTGCTTTTGTGTTATCCAGAGGGAAAGGGACGCCTGTTTG
TGGGCAGAGCACAACTGTTCTTTATGTGCGATGCAGTCGCTGCCACAAGTAGGAAAATA
TGGAGTCAGCTGCACCGTAGCACCTTCACTATCCCCAGTCACTGGAATCACCCTGAAGC
TCGCGTTTGTCTTCTGGTGTGGCTCAAACACCTGAGAAAACCCGATGATAGTGTGTG
35 TTGGATAAGAGCCATGTATCTGTTTCCAGGGGAGTGGTTGTGAAGTTCACCTTCTACTGT
CTCCTCATTCTTCTTACAAGCAGTGATGTTCCGATCCCACAGGCTTCCGGCCTTGACACA
CTNTNTTTTATATTTTATTATGTGGTCTAGGCAGCCTGGTGAGGTGAAATTCACAGACAT
GGAAGGGCCATCTTCATTCATATTTGCATTAGGAATATTATGGGCCCAATGAAATAGAC
TGTGTTTCACTCTACAGGGGAAGCCGATATAGGAAAATGTCCATTTACCACCAGAGGGTC
40 TGGTCTGAGTCTTGAAGGCCTTTTGTGTTATTGCACCTTACACAGCTGTTAGACTGGGAA
GTTGCTTTTGGCCCGCACACAAATCTTGTGGGCCTTCAACAGCGGATGCTGCCATTTGCC
CCGAAGTCCCAGCTCAATTCATTAAAAATTGAATAGGCCCTTGTGGCAACCCTAGTTG

GTACAGGGTTTTACTTGGGGGGCCCCCTCTAAGTTTCCCCGGGATATAAACAAAGTGTGG

BX453537

5 TTATGGGGGGCAGTAGTGTGGTGGTAGAAAAGGAAGTCTTCTTGATCCTTTTCGTGCCTCC
ACATTAGATAGATCCCTGCCACCAGCACCCATGTGGCCACCAGCAGAGACAGCAGGAGGA
GAGGCAGCCAGCCTCCCGGCTTGCTTTTGTGTATCCAGAGGGAAAGGGACGCCTGTTT
GTGGGCAGAGCACAACTGTTCCCTTATGTCTGGATGCAGTCGCTGCCACAAGTAGGAAAAT
ATGGAGTCAGCTGCACCGTAGCACCTTCACTATCCCCAGTCACTGGAATCACCACTGAAG
CTCGCGTTTGTCTTCTGCTGTGGCTCAAACACCTGAGAAAACCCGATGATAGTGCTGT
10 GTTGGATAAGAGCCATGTATCTGTTTCCCAGGGGAGTGTTGTGAAGTTCACTTCTACTG
TCTCCTCATTCTTCTTACAAGCAGTGATGTTTCGGATCCCACAGGCTTCCGGCCTTGACAC
ACTTTTTTTTATATTTTATTATGTGGTCTAGGCAGCCTGGTGAGGTGAAATTCACAGACA
TGGAAGGGCCATCTTCATTCATATTTGCATTAGGAATATTATGGGCCCAATGAAATAGA
CTGTGTTTCACTCTACAGGGAAGCCGATATAGGAAAATGTCCATTTACCACCAGAGGGTC
15 TGGTCTGAGTCTGGAAGGCCTCTGTGTAATTGCACCTCACACAGCTGTAGGACTGGGAGT
TGCTTTTGGCCGTACACAAATCTTGTTGGCCTTCAACAAGCGGATGCTGGCATCTGGCGG
GGGTACCCAGCTTACATTCATCAAAATTGAATAGTCCCCTTGTTGCAACACTAGTTTGTA
AACAGGTTCTACTCCGGGGTCCCCTCAGTCTCCCGG

20 **AV728945**

CAAAATATGAATGAAGATGGCCCTTCCATGTCTGTGAATTTACCTCACCAGGCTGCCTAG
ACCACATAATGAAATATAAAAAAAGTGTGTCAAGGCCGGAAGCCTGTGGGATCCGAACA
TCACTGCTTGTAAGAAGAATGAGGAGACAGTAGAAGTGAACCTCACAACCACTCCCCTGG
GAAACAGATACATGGCTCTTATCCAACACAGCACTATCATCGGGTTTTCTCAGGTGTTTG
25 AGCCACACCAGAAGAAACAAACGCGAGCTTCAGTGGTGATTCCAGTGACTGGGGATAGTG
AAGGTGCTACGGTGCAACTGACTCCATATTTTCTACTTGTGGCAGCGACTGCATCCGAC
ATAAAGGAACAGTTGTGCTCTGCCACAAACAGGCGTCCCTTTCCCTCTGGATAACAAC

AV728939

30 GCAAATATGAATGAAGATGGCCCTTCCATGTCTGTGAATTTACCTCACCAGGCTGCCTA
GACCACATAATGAAATATAAAAAAAGTGTGTCAAGGCCGGAAGCCTGTGGGATCCGAAC
ATCACTGCTTGTAAGAAGAATGAGGAGACAGTAGAAGTGAACCTCACAACCACTCCCCTG
GGAAACAGATACATGGCTCTTATCCAACACAGCACTATCATCGGGTTTTCTCAGGTGTTT
GAGCCACACCAGAAGAAACAAACGCGAGCTTCAGTGGTGATTCCAGTGACTGGGGATAGT
35 GAAGGTGCTACGGTGCACTGACTCCATATTTTCTACTTGTGGCAGCGACTGCATCCGA
CATAAAGGAACAGTTGTGCTCTGCCACAAACAGGCGTCCCTTTCCCTCTGGATAACAAC

AV727345

40 GCAAATATGAATGAAGATGGCCCTTCCATGTCTGTGAATTTACCTCACCAGGCTGCCTA
GACCACATAATGAAATATAAAAAAAGTGTGTCAAGGCCGGAAGCCTGTGGGATCCGAAC

ATCACTGCTTGTAAGAAGAATGAGGAGACAGTAGAAGTGAAC TTCACAACCACTCCCCTG
GGAAACAGATACATGGCTCTTATCCAACACAGCACTATCATCGGGTTTTCTCAGGTGTTT
GAGCCACACCAGAAGAAACAAACGCGAGCTTCAGTGGTGATTCCAGTGACTGGGGATAGT
GAAGGTGCTACGGTGCAGCTGACTCCATATTTTCTACTTGTGGCAGCGACTGCATCCGA
5 CATAAAGGAACAGTTGTGCTCTGCCCCACAAACAGGCGTCCCTTTCCCTCTGGATAACAAC
AAAAGCAAGCCGGGAGGCTGGCTGCCTCTCCTCCTGCTGTCTCTGCTGGTGGCCACATGG
GTGCTGGTGGCAGGGATCTATCTAATGTGGAGGCACGAAAGGATCAAGAAGACTTCCTTT
TTTACCACCACACTACTGTCTCCCATTAAGATCTTGTGGTTTATCCATCTGAAATATTG
TTCCATTACACATATTGGTACCTAACTGAAATTCTTTAAACCATTGCAAATTGAGGTCA
10 CTCTTGAAAGGGCGTG

Sequences identified as those of CACNA1D cluster

BM128550

15 CGGCTCCTACCTTTTGCCCGATCCCCTTCCCCATTCCGCCCCCGCCCCAACGCAGTGCAC
AGTGCCCTGCACACAGTAGTCGCTCAATAAATGTTCTGGATGATGATGATGATGATGAT
GAAAAAATGCAGCATCAACGGCAGCAGCAAGCGGACCACGCGAACGAGGCAAACCTATGC
AAGAGGCACCAGACTTCCTCTTTCTGGTGAAGGACCAACTTCTCAGCTGAATAGCTCCAA
GCAAACCTGTCCTGTCTTGGCAAGCTGCAATCGATGCTGCTAGACAGGCCAAGGCTGCCCCA
20 AACTATGAGCACCTCTGCACCCCCACCTGTAGGATCTCTCTCCCAAAGAAAACGTCAGCA
ATACGCCAAGAGCAAAAAACAGGGTAACTCGTCCAACAGCCGACCTGCCCGCGCCCTTTT
CTGTTTATCACTCAATAACCCCATCCGAAGAGCCTGCATTAGTATAGTGGAATGGAAACA
TTTGACATATTTATATTATTGGCTATTTTTTTGCCAAT

25 BI755471

GAATATGACCCTGAGGCAAAGGGAAGGATAAACACCTTGATGTGGTCACTCTGCTTCGAC
GCATCCAGCCTCCCCTGGGGTTTGGGAAGTTATGTCCACACAGGGTAGCGTGCAAGAGAT
TAGTTGCCATGAACATGCCTCTCAACAGTGACGGGACAGTCATGTTTAATGCAACCCTGT
TTGCTTTGGTTTGAACGGCTCTTAAGATCAAGACCGAAGGGAACCTGGAGCAAGCTAATG
30 AAGAACTTCGGGCTGTGATAAAGAAAATTTGGAAGAAAACAGCATGAAATTACTTGACC
AAGTTGTCCCTCCAGCTGGTGATGATGAGGTAACCGTGGGGAAGTTCTATGCCACTTTCC
TGATACAGGACTACTTTAGGAAATTCAGAAACGGAAAGAACCAAGGACTGGTGGGAAAGT
ACCCTGCGAAGAACACCACAATTGCCCTACAGGCGGGATTAAGGACACTGCATGACATTG
GGCCAGAAATCCGGCGTGCTATATCGTGTGATTTGCAAGATGACGAGCCTGAGGAAACAA
35 AACGAGAAGAAGAAGATGATGTGTTCAAAGAAATGGTGCCCTGCTTGGAACCATGTCA
ATCATGTTAATAGTGATAGGAGAGATTCCCTTCAGCAGACCAATAGCACCACCGTCCCCCT
GCATTGTCCAAAGGCCTTCAATTCCACCTGCAAGTGATACTGAGAAACCGCTGTTTCCTC
CAGCAGGAAATTCGGGGTGTGATAACCATCATAACCATTAATTCCATAGGAAAGCAAGGT
TCCCATTCAACAATGCCAGTCTCGAATAGTGCCAATATGTCCAAAGCTTGCCATGGTAA
40 GCGGGCCAGCATTGGGAACC

BQ549084

GCACGAGATTAATTAGACTTTTGTATAAGAGATGTCATGCCTCAAGAAAGCCATAAACCT
GGTAGGAACAGGTCCCAAGCGGTTGAGCCTGGCAGAGTACCATGCGCTCGGCCCCAGCTG
CAGGAAACAGCAGGCCCCGCCCTCTCACAGAGGATGGGTGAGGAGGCCAGACCTGCCCTG
5 CCCCATTGTCCAGATGGGCACTGCTGTGGAGTCTGCTTCTCCCATGTACCAGGGCACCAG
GCCACCCAACTGAAGGCATGGCGGCGGGGTGCAGGGGAAAGTTAAAGGTGATGACGATC
ATCACACCTGTGTCGTTACCTCAGCCATCGGTCTAGCATATCAGTCACTGGGCCCCAACAT
ATCCATTTTTTAAACCCTTTCCCACAAATACACTGCGTCCTGGTTCCTGTTTAGCTGTTCT
GAAATACGGTGTGTAAGTAAGTCAGAACCCAGCTACCAGTGATTATTGCGAGGGCAATGG
10 GACCTCATAAATAAG

BQ549571

TTTTTTTTTTTTTTTTTTTTTAGTGGGGAACCTACAATTATTAGGACCCATGGATATTGCTGCA
GTTCAAATACAATACAGTAATTACAAAATATAGACCATCTCTTTACAAATACAAATTATA
15 GTATATTACAAGTCATGTACAGTAAATCTATAATTTTAAACAAACTAGTGTATCTAAGTT
TACCTGGTTGCGAGTGCATTATTATTCCAGTTTACAGTTGCCCTTAGCGTGACAGTCAGA
AACCGACCATCGGAGTGATATTCTCTTATGTAAACTGGCGTCACATCACAGAAAACCTTA
TTTATGAGGTCCCATTTGCCCTCGCAATAATCACTGGTAGCTGGGTTCTGACTTACTTACA
CACCGTATTTTCAGAACAGCTAAACAGGAACCAGGACGCAGTGTATTTGTGGGAAAGGGTT
20 TAAAAATGGATATGTTGGGCCCAGTGACTGATATGCTAGACCGATGGCTGAGGTAACGAC
ACAGGTGTGATGATCGTCATCACCTTTAACTTTCCCCTGCACCCCGCCGCCATGCCTTCC
AGTTGGGTGGGCCTGGT

AI693324

CTCTGAGCACTACAATCAGCCAGATTGGTTGACACAGATTCAAGATATTGCCAACAAAGT
CCTCTTGGCTCTGTTCACCTGCGAGATGCTGGTAAAAATGTACAGCTTGGGCCTCCAAGC
ATACTCTTGTCTCTTTACAACCGGTTTGATTGCTTCGTGGTGTGTGGTGGAATCACTGA
GACGATCTTGGTGGAACCTGGAAATCATGTCTCCCCTGGGGATCTCTGTGTTTCGGTGTGT
GCGCCTCTTAAGAATCTTCAAAGTGACCAGGCACTGGACTTCCCTGAGCAACTTAGTGGC
30 ATCCTTATTAAACTCCATGAAGTCCATCGCTTCGCTGTTGCTTCTGCTTTTTCTCTTCAT
TATCATCTTTTCCTTGCTTGGGATGCAGCTGTTTGGCGGCAAGTTTAATTTTGATG

R25307

ACCAGCAGACCTGACTGTCCCCAGCAGCTTCCGGAACAAAAACAGCGACAAGAGAGGAGT
35 GCGGACAGTTGGTGAGGCAGTCCTGATATCCGAAGCTTGGGACGCTATGCAAGGGACCC
AAAATTTGTGTCAGCAACAAAACACGAAATCGCTGATGCCTGTGACCTCACCATCGACGA
GATGGAGAGTGCAGCCAGCACCCCTGCTTAATGGGAACGTGCGTCCCCGAGCCAACGGGGA
TGTGGGCCCCCTCTCACACCGGCAGACTATGAGCTACAGGACTTTGGTCCTGGGCTTACA
GCGACGAAGAGCCAGACCCTGGGGAGGGATTGAGGGAGGACCTGGGCGGATGAATTGATT
40 TTGCNTCACCACCTTTGTTAGGCCCCCAGGCGAGGGGCAAG

R46658

TTTTTTTTTTNTTTTTTTTTTTGTGGAAAGATGATAGGTTTATAGTGACTCAAAATATTT
TAGAAAAATTTCTGTAGTGTCAAGTTCTTTCAAACCTAAAATTTTAACCCCAGAGGATTT
TCGCTGAATAAATGAGAATTGGCTCTATTTCTTCTACTTCTGGATAGCCCGNGTAAAAAT
5 ACTAAT

H29256

TTTTTTTTTTTTTTTTTTTTTTGTGGAAAGATGATAGGTTTATAGNGACTCAAAATATTTT
AGAAAAATTTCTGTAGTGTCAAGTTCTTTCAAACCTAAAATTTTAACCCCAGAGGATTTT
10 CGCTGAATAAATGAGAATTGGCTCTATTTCTTCTACTTCTGGATAGCCCGAGTAAAAATA
CTAATAATTTCTAGATTTTGTAGTGGGGAACCTACAATTATTAGGACCCATGGATATTGCTGC
AGTTCAAATACAATACAGTAATTACAAAATATAGACCATCTCTTTACAAATACAANTTAT
AGNATATTACAAGTCATGTACAGTAAATCTATAATTTTGGACAANCTAGTGTATCTAAGT
TTACCNGGGGTGCGAGTGCCTTATTNTTCCNGTTTACAGTTGCCCTTAGCGTGACAGTCN
15 GGAACCGNCCTTC

H29339

GCCTGACTGTCCCCAGCAGCTTCCGGAACAAAAACAGCGACAAGCAGAGGAGTGCGGACA
NTTTGGTGGAGGCAGTCCTGATATCCGAAGCTTGGGACGCTATGCAAGGGACCCAAAATT
20 TGTGTCAGCAACAAACACGAAATCGCTGATGCCTGTGACCTCACCATCGACGAGATGGA
GAGTGCAGCCAGCACCCCTGCTTAATGGGAACGTGCGTCCCCGAGCCAACGGGGATGTGGG
CCCCCTCTCACACCGGCAGACTATGAGCTACAGGACTTTGGTCTCTGGGCTACAGCGACGA
AGAGCCAGACCCTGGGAGGGATGAGGAGGAC

BG716371

AGCGGTCGTAATAATGTAGTTCCCCACTAAAATCTAGAAATTATTAGTATTTTTACTCGG
GCTATCCAGAAGTAGAAGAAATAGAGCAAATTTCTCATTTATTCAGCGAAAATCCTCTGGG
GTTAAAATTTTAAGTTGAAAGAACTTGACACTACAGAAATTTTCTAAAATATTTGAGTC
ACTATAAACCTATCATCTTTCCACAAGATATACCAGATGACTATTGCAGTCTTCTCTTGG
30 GCAAGAGTTCCATGATTTGATACTGTACCTTGGATCCACCATGGGTGCAACTGTCTTGGT
TTGTTGTTGACTTGAACCACCCTCTGGTAAGTAAGTGAATTACAGAGCAGGTCTAGCTGG
CTGCTCTGCCCCCTTGGGTATCCATAGTTACGGTTTTCTCTGTGGCCACCCAGGTGTTTT
TGCATCGCTGGTGCAGAAATGCACAGGTGGATGAGATATAGCTGCTCTTGTCCTCTGGGG
ACTGGTGGTGTGCTTAAGAAATAAGGGGTGCTGGGGACAGAGGAGCAACGTGGTGATCT
35 ATAGGATTGGAGTGTGCGGGTCTGTACAAATCGTATTGTTGCCTTTTACAAAACCTGTGTA
CTGTATGTTCTCTTTGAGGGCTTTTGTATGCAATTGAATGAGG

AI537488

TTTTCTTGTGGAAAGATGATAGGTTTATAGTGACTCAAAATATTTTAGAAAAATTTCTGT
40 AGTGTCAAGTTCTTTCAAACCTAAAATTTTAACCCCAGAGGATTTTCGCTGAATAAATGA

GAATTGGCTCTATTTCTTCTACTTCTGGATAGCCCGAGTAAAAATACTAATAATTTCTAG
ATTTTAGTGGGGAAC TACAATTATTAGGACCCATGGATATTGCTGCAGTTCAAATACAAT
ACAGTAATTACAAAATATAGACCATCTCTTTACAAATACAAATTATAGTATATTACAAGT
CATGTACAGTAAATCTATAATTTTAAACAAACCTGTGTATCTAAGTTTACCTGGTTG

5

AA458692

GACAAATAAAGCAATTATAAATGTATCTCACTTTAGAACAGACAAAAAAGGGCATGCTA
TGGAAATTGTTTAAATCTCAAGCAACAATGCTGATTAATTTCTGGTCAATAATCGTTCTA
TAGTTCTCCTTCATGAAGCCTGGTGAGGTTCCAGGAAACAGCTTGATTTGGGAAGCCTCA
GCAGAAAAGAAAGCATCTCAGAGGACACATAAAATGTCTGGCAACCCCTCTTGGCGGCC
TCATCCAGCAAAGCTTGTGTGGTCTTGGCAACTGTCCTCAGGACTCTGCTTTCAAGATGA
AAGAGGTGTAGCTTACCCGCTCAATACACCAAGTACAAGATTTAGTACGAAAAATGACCC
AAAGATGACGAGACTGACAAGATACACCCAGGGCAATTCCAATCCCATAGCATCATTCAT

10

15

AI393327

TTTATATTATTCACCACTTTGTTATGAAGACCTTACAAACCTCTTCTTAAGACATTCTTA
CTCTGATCCAGGCAAAAACACTTCAAGGTTTGTAATGACTCTTTCCTGACATAAATCCT
TTTTTATTAAAATGCAAAATGTTCTTCAGAATAAACTGTGTAATAATTTTATACTTGG
GAGTGCTCCTTGACAGAGCTGTCATTTGCCAGTGAGAGCCTCCGACAGGGCAGGTACTG
TGCCAGGGCAGCTCTGAAATTATGGATATTCTTATCCTCCTGGTTCCTTCGGTGCCAATG
GTAACCTAATACCAGCCGCAGGGAGCGCCATTTCTCCTAAAGGGCTACACCACTGTCAAC
ATTATCCTGGACTCTGTGTCTCTCTGTGGGTCTTGTGGCATCACATCAGGCCAAAAT
TGCCAGACCAGGACCCTAAGTGTCTGATAGAGGCGATGATCTTTT

20

25

AI520947

TTTTTTTTTTTTTTTTTTTTTTTTCTTACAAAGAAAAATTTAATATTCGATGAGAGGTTGAAC
CAGGCTTAAAGCAAACATACTAGGAAATGGGGCAGCCTGTAAGAATGCCAGTTTGTAAGT
ACTGACTTTGGAAAAGATCATCGCCTCTATCAGACACTTAGGGTCCTGGTCTGGCAATTT
TGGCCTGATGTGATGCCACAAGACCCAACAGAGAGAGACACAGAGTCCAGGATAATGTTG
ACAGGGGGTAGCCCTTTAGGAGAAATGGCGCTCCCTGCGGCTGGTATTAGGTTACCATTG
GCACCGAAGGAACCAGGAGGATAAGAATAT

30

AI248998

TGTAAATAACAAACACCACTTTGTTATGAAGACCTTACAAACCTCTTCTTAAGACATTCT
TACTCTGATCCAGGCAAAAACACTTCAAGGTTTGTAATGACTCTTTCCTGACATAAATC
CTTTTTTATTAAAATGCAAAATGTTCTTCAGAATAAACTGTGTAATAATTTTATACTT
GGGAGTGCTCCTTGACAGAGCTGTCATTTGCCAGTGAGAGCCTCCGACGGGGCAGGTAC
TGTGCCAGGGCAGCTCTGAAATTATGGATATTCTTATCCTCCTGGTTCCTTCGGTGCCAA
TGGAACCTAATACCAGCCGCAGGGAGCGCCATTTCTCCTAAAGGGCTACACCACTGTCA
ACATTATCCTGGACTCTGTGTCTCTCTGTGGGTCTTGTGGCATCACATCAGGCCAAA

35

40

ATTGCCAGACCAGGACCCTAAGTGTCTGATAGAGGCGATGATCTTTTCCAAAGTCAGTAC
TTACAAACTGGCATTCTTACAG

AI075844

5 TTTTTTTTTTTGTAAATAACAAACACCACTTTGTTATGAAGACCTTACAAACCTCTTCTT
AAGACATTCTTACTCTGATCCAGGCAAAAACACTTCAAGGTTTGTAATGACTCTTTCCT
GACATAAATCCTTTTTTTATTAAAATGCAAAATGTTCTTCAGAATAAAACTGTGTAATAAT
TTTTATACTTGGGAGTGCTCCTTGCACAGAGCTGTCATTTGCCAGTGAGAGCCTCCGACG
GGGCAGGTACTGTGCCAGGGCAGCTCTGAAATTATGGATATTCTTATCCTCCTGGTTCCCT
10 TCGGTGCCAATGGTAACCTAATACCAGCCGCAGGGAGCGCCATTTCTCCTAAAGGGCTAC
ACCACTGTCAACATTATCCTGGACTCTGTGTCTCTCTCTGTTGGGTCTTG

AI869807

GTAAATAACAAACACCACTTTGTTATGAAGACCTTACAAACCTCTTCTTAAGACATTCTT
15 ACTCTGATCCAGGCAAAAACACTTCAAGGTTTGTAATGACTCTTTCCTGACATAAATCC
TTTTTTATTAAAATGCAAAATGTTCTTCAGAATAAAACTGTGTAATAATTTTTTATACTTG
GGAGTGCTCCTTGCACAGAGCTGTCATTTGCCAGTGAGAGCCTCCGACGGGCAGGTACTG
TGCCAGGGCAGCTCTGAAATATGGATATTCTTACCTCCTGGTTCTTTCGGTGCAAATGGT
AACCTAATACCAGCCGCAGGGAGCGCCATTTCT

20

AI869800

GTAAATAACAAACACCACTTTGTTATGAAGACCTTACAAACCTCTTCTTAAGACATTCTT
ACTCTGATCCAGGCAAAAACACTTCAAGGTTTGTAATGACTCTTTCCTGACATAAATCC
TTTTTTATTAAAATGCAAAATGTTCTTCAGAATAAAACTGTGTAATAATTTTTTATACTTG
25 GGAGTGCTCCTTGCACAGAGCTGTCATTTGCCAGTGAGAGCCTCCGACGGNGCAGGTACT
GTGCCAGGGCAGCTCTGAATTATGGATATTCTTATCCTCCTG

25

AI243110

TTTTTCTTACAAAGAAAAATTTAATATTTCGATGAGAGGTTGAACCAGGCTTAAAGCAGAC
30 ATACTAGGAAATGGTGCAGCCTGTAAGAATGCCAGTTTGTAAGTACTGACTTTGGAAAAG
ATCATCGCCTCTATCAGACACTTAGGGTCTGGTCTGGCAATTTTGGCCTGATGTGATGC
CACAAGACCCAACAGAGAGAGACACAGAGTCCAGGATAATGTTGACAGTGGTGTAGCCCT
TTAGGAGAAATGGCGCTCCCTGCGGCTGGTATTAGGTTACCATTGGCACCGAAGGAACCA
GGAGGATAAGAATATCCATAATTTTACAGAGCTGCCCTGGCACAGTACCTGCCCCGTCGGAG
35 GCTCTCACTGGCAAATGACAGCTCTGTGCAAGGAGCACTC

35

AI955764

TTATCTTGTGGAAAGATGATAGGTTTATAGTGACTCAAAATATTTTAGAAAAATTTCTGT
AGTGTCAAGTTCTTTCAAACTTAAAATTTTAACCCAGAGGATTTTCGCTGAATAAATGA

GAATTGGCTCTATTTCTTCTACTTCTGGATAGCCCGAGTAAAAATACTAATAATTTCTAG
ATTTTAGTGGGGAAC TACAATTATTAGGACCCATGGATATTGCTGCAGTTCAAATACAAT
ACAGTAATTACAAAATATAGACCATCTCTTTACAAATACAAATTATAGGATATTACAAGG
CATGTACAGTAAATCTATAATTTTAAACAAACTAGTGTATCTAAGTTTACCTGGTTGCGA
5 GTGCATTATTATTCCAGTTTACAGTTGCCCTTAGCGTGACAGTCAGAAACCGACCATCGG
AGTGATATTCTCTTATGTAAACTGGCGTCACATCACAGAAAACCTTATTTATGAGGTCCC
AT

AA192669

10 GCCCTCACAGCCCACCACGCCTGGCCTTCGCCCAATTCTGAAACTTCGTAGGATAGAGCT
GGAAAGTGCCACATGGTGAAGCGAGATCCAGCTGTCTGGGTGGATGTCGGAGTCCATAGG
CTGAGCAGAGATGGTTCTTAGTGAGGTTCTCGCTGCCAGTTGACGGTGAAATCATAGCTG
CCATTTACATTTTGTGAGATTATGAAAAACATAAGACTAAAGAACTAAATGTGTATTTC
CTGTGGACACAAAAATGTGTGTTTTTTCAGATGGGGAGGGGACCAAAAAGGAAAAACATTT
15 CATCTTAAAACTTTCCTAAGACAAAGGAAAACAAAAAACCATGCTCCTACAACCTTCAAAT
TTTTCTTACCAAAGAAAAATTTAATATTTCGATGAGAGGTTGAACCAGGCTTAAAGCAGAC
ATACTAGGGAATGGGTGCAGCCTGTAAGAATGCCAGTTT

AA192157

20 GTAAATAACAAACACCACTTTGTTATGAAGACCTTACAAACCTCTTCTTAAGACATTCTT
ACTCTGATCCAGGCAAAAACACTTCAAGGTTTGTAATGACTCTTTCCTGACATAAATCC
TTTTTTATTAAATGCAAAATGTTCTTCAGAATAAACTGTGTAATAATTTTTTATACTTG
GGAGTGCTCCTTGCACAGAGCTGTCATTTGCCAGTGAGAGCCTCCGACGGGGCAGGTACT
GTGCCAGGGCAGCTCTGAAATTATGGATATTCTTATCCTCCTGGTTCCTTCGGTGCCAAAT
25 GGTAACCTAATACCAGCCGCAGGAGCGCCATTTCTCCTAAAGGGCTACACCACTGTCAAC
ATTATCCTGGGACTCTGTGTCTCTCTCTGTTGGGTCTTGTGGCATCACATCAGGCCAAAA
TTGGCCAGACCAGGACCCCAAGTGGTCTGATAGAAGGCGATGATCTTTTCCAAAGTCAGT
ACTTACA

AI361691

30 GTTTAAAATTATAGATTTACTGTACATGACTTGTAATATACTATAATTTGTATTTGTAAA
GAGATGGTCTATATTTTGTAATTACTGTATTGTATTTGAACTGCAGCAATATCCATGGGT
CCTAATAATTGTAGTTCCCCACTAAAATCTAGAAATTATTAGTATTTTTTACTCGGGCTAT
CCAGAAGTAGAAGAAATAGAGCCAATTCTCATTTATTTCAGCGAAAATCCTCTGGGGTTAA
35 AATTTTAAGTTTGAAAGAACTTGACACTACAGAAATTTTTCTAAAATATTTTGAGTCACT
ATAAACCTATCATCTTTCCACAAGATATACCAGATGACTATTTGCAGTCTTTTCTTTGGG
CAAGAGTTCCATGATTTTGATACTGTACCTTTGGATCCACCATGGGTTGCAACTGTCTTT
GGTTTGTGTTGTTGACTTGAACCA

AI914244

TTCGCTGAATAAATGAGAATTGGCTCTATTTCTTCTACTTCTGGATAGCCCGAGTAAAAA
TACTAATAATTTCTAGATTTTAGTGGGGAAC TACAATTATTAGGACCCATGGATATTGCT
GCAGTTCAAATACAATACAGTAATTACAAAATATAGACCATCTCTTTACAAATACAAATT
ATAGTATATTACAAGTCATGTACAGTAAATCTATAATTTTAAACAAACTAGTGTATCTAA
5 GTTTACCTGGTTGCGAGTGCATTATTATTCCAGTTTACAGTTGCCCTTAGCGTGACAGTC
AGAAACCGACCAT

AW008769

TTTTATCTTGTGGAAAGATGATAGGTTTATAGTGACTCAAAATATTTTAGAAAAATTTCT
10 GTAGTGTCAAGTTCTTTCAAAC TTAATAACCCAGAGGATTTTCGCTGAATAAAT
GAGAATTGGCTCTATTTCTTCTACTTCTGGATAGCCCGAGTAAAAATACTAATAATTTCT
AGATTTTAGTGGGGAAC TACAATTATTAGGACCCATGGATATTGCTGCAGTTCAAATACA
ATACAGTAATTACAAAATATAGACCATCTCTTTACAAATACAAATTATAGTATATTACAA
GTCATGTACAGTAAATCTATAATTTTAAACAAACTAGTGTATCTAAGTTTACCTGGTTGC
15 GAGTGCATTATTATTCCAGTTTACAGTTGCCCTTAGCGTGACAGTCAGAAACC

AW008794

TTTTATCTTGTGGAAAGATGATAGGTTTATAGTGACTCAAAATATTTTAGAAAAATTTCT
GTAGTGTCAAGTTCTTTCAAAC TTAATAACCCAGAGGATTTTCGCTGAATAAAT
20 GAGAATTGGCTCTATTTCTTCTACTTCTGGATAGCCCGAGTAAAAATACTAATAATTTCT
AGATTTTAGTGGGGAAC TACAATTATTAGGACCCATGGATATTGCTGCAGTTCAAATACA
ATACAGTAATTACAAAATATAGACCATCTCTTTACAAATACAAATTATAGTATATTACAA
GTCATGTACAGTAAATCTATAATTTTAAACAAACTAGTGTATCTAAGTTTACCTGGTTGC
GAGTGCATTATTATTCCAGTTTACAGTTGCCCTTAGCGTGACAGTCAGAAACCGACCATC
25 GGAGTGATATTCTCTTATGTAAACTGGCGTCACATCACAGAAACCTTATTTATTT

AA877582

TTTTTTTTTTAGAGCCAATTCTCATTTATTTCAGCGAAAATCCTCTGGGGTTAAAATTTTA
AGTTTGAAAGAACTTGACACTACAGAAATTTTCTAAAATATTTTGAGTCACTATAAACC
30 TATCATCTTTCCACAAGATATACCAGATGACTATTTGCAGTCTTTTCTTTGGGCAAGAGT
TCCATGATTTTGATACTGTACCTTTGGATCCACCATGGGTGCAACTGTCTTTGGTTTTG
TTTGTTTGACTTGAACCACCTCTGGTAAGTAAGTGAATTACAGAGCAGGTCCAGCTGGC
TGCTCTGCCCCCTTGGGTATCCATAGTTACGGTTTTCTCTGTGGCCCAACCCAGGGTGT
TTGCATCGCTGGTGCAGAAATGCACAGGTGGATGAGATATAGCTGC
35

AI051972

TTTTTTTTTTGTAAATAACAAACACCACTTTGTTATGAAGACCTTACAAACCTCTTCTTAA
GACATTCTTACTCTGATCCAGGCAAAAACACTTCAAGGTTTGTAAATGACTCTTTCCTGA
CATAAATCCTTTTTTTATTAAAATGCAAAATGTTCTTCAGAATAAAACTGTGTAATAATTT
40 TTATACTTGGGAGTGCTCCTTGACACAGAGCTGTCATTTGCCAGTGAGAGCCTCCGACAGG

GCAGGTACTGTGCCAGGGCAGCTCTGAAATTATGGATATTCTTATCCTCCTGGTTCCTTC
GGTGCCAATGGTAACCTAATACCAGCCGAGGGAGCGCCATTTCTCCTAAAGGGCTACAC
CACTGTCAACATTATCCTGGACTCTGTGTCTCTCTGTGAGTCTTGTGGCATCACATC
AGGCCAAAATTGCCAGACCAGGACCCTAAGTGTCTGATAGAGGCGATGATCTT

5

AI017959

TTTAGAGCCAATTCTCATTTATTCAGCGAAAATCCTCTGGGGTTAAAATTTTAAGTTTGA
AAGAACTTGACACTACAGAAATTTTTCTAAAATATTTTGAGTCACTATAAACCTATCATC
TTTCCACAAGATATAACCAGATGACTATTTGCAGTCTTTTCTTTGGGCAAGAGTTCCATGA
TTTTGATACTGTACCTTTGGATCCACCATGGGTTGCAACTGTCTTTGGTTTTGTGTGTT
GACTTGAACCAACCTCTGGTAAGTAAGTGAATTACAGAGCAGGTCCAGCTGGCTGCTCTG
CCCCTTGGGTATCCATAGTTACGGTTTTCTGTGGCCCAACCCAGGGTGTTTTTTGCATC
GCTGGTGCAGAAATGCACAGGTGGATGAGATATAGCTGCTCTTGTCTCTGGGGACTGGT
GGTGCTGCTTAAGAAATAAGGGGTGCTGGGGACAGAGGAGCAA

10

15

N79331

TTTTTTTTTTTTTTGTAAATAACAAACACCACTTTGTTATGAAGACCTTACAAACCTCTT
CTTAAGACATTCTTACTCTGATCCAGGCAAAAACACTTCAAGGTTTGTAAATGACTCTTT
CCTGACATAAATCCTTTTTTG

20

N62240.1

ACAAAGAAAAATTTAATATTCGATGAGAGGTTGAACCAGGCTTAAAGCAGACATACTAGG
AAATGGTGCAGCCTGTAAGAATGCCAGTTTGTAAGTACTGACTTTGGAAAAGATCATCGC
CTCTATCAGACACTTAGGGTCTGGTCTGGCAATTTTGGCCTGATGTGATGCCACAAGAC
CCAACAGAGAGAGACACAGAGTCCAGGNTAATATTGACAGNAGGTGGANGCCCCCT

25

AI240933

TTTTTTTTTTTTTTTTTTTTTGGTCCAAAATTTTAAATAGTATACAGACAACCTGTTAATT
TTTTTTTTTTTTTTTTTTTTTGGAAATAACAAACACCACTTTGTTATGAAGACCTTACAAACC
TCTTCTTAAGACATTCTTACTCTGATCCAGGCAAAAACACTTCAAGGTTTGGAAATGACT
CTTTCCTGACATAAATCCTTTTTTATTAATAATGCAAAAGGTTCTTCAGAATAAACTGTG
TAATAATTTTTTATACTTGGGAGTGCTCCTTGCACAGAGCTGTCATTTGCCAG

30

AI015031

TTTTTCTTACAAAGAAAAATTTAATATTCGATGAGAGGTTGAACCAGGCTTAAAGCAGAC
ATACTAGGAAATGGTGCAGCCTGTAAGAATGCCAGTTTGTAAGTACTGACTTTGGAAAAG
ATCATCGCCTCTATCAGACACTTAGGGTCTGGTCTGGCAATTTTGGCCTGATGTGATGC
CACAAGACCCAACAGAGAGAGACACAGAGTCCAGGATAATGTTGACAGTGGTGTAGCCCT
TTAGGAGAAATGGCGCTCCCTGCGGCTGGTATTAGGTTACCATTGGCACCGAAGAGACCA

35

GGAGGATAAGAATATCCATAATTTTCAGAGCTGCCCTGGCACAGTACCTGCCCCGTCGGAG
GCTCTCACTGGCAAATGACAGCTCTGTGCAAGGAGCACTCCCAAGTATAAAAATTATTAC
ACAGTTTTATTCTG

5 AI290994

TAAATAACAAACACCACTTTGTTATGAAGACCTTACAAACCTCTTCTTAAGACATTCTTA
CTCTGATCCAGGCAAAAACACTTCAAGGTTTGTAAATGACTCTTTCCTGACATAAATCCT
TTTTTATTAAAATGCAAAATGTTCTTCAGAATAAAACTGTGTAATAATTTTTTATACTTGG
GAGTGCTCCTTGACAGAGCTGTCATTTGCCAGTGAGAGCCTCCGACGGGGCAGGTACTG
10 TGCCAGGGCAGCTCTGAAATTATGGATATTCTTATCCTCCTGGTTCCTTCGGTGCCAATG
GTAACCTAATACCAGCCGCAGGGAGCGCCATTTCTCCTAAAGGGCTACACCACTGTCAAC
ATTATCCTGGACTCTGTGTCTCTCTGTGGGTCTTGTGGCATCACATCAGGCCAAAAT
TGCCAGACCAGGACCCTAAGTGTCTGATAGA

15 AA861160

TTTGTAATAACAAACACCACTTTGTTATGAAGACCTTACAAACCTCTTCTTAAGACATT
CTTACTCTGATCCAGGCAAAAACACTTCAAGGTTTGTAAATGACTCTTTCCTGACATAAA
TCCTTTTTTATTAAAATGCAAAATGTTCTTCAGAATAAAACTGTGTAATAATTTTTTATAC
TTGGGAGTGCTCCTTGACAGAGCTGTCATTTGCCAGTGAGAGCCTCCGAC

20

 AA915941

TTTGTAATAACAAACACCACTTTGTTATGAAGACCTTACAAACCTCTTCTTAAGACATT
TTACTCTGATCCAGGCAAAAACACTTCAAGGTTTGTAAATGACTCTTTCCTGACATAAAT
CCTTTTTTTATTAAAATGCAAAATGTTCTTCAGAATAAAACTGTGTAATAATTTTTTATACT
25 TGGGAGTGCTCCTTGACAGAGCTGTCATTTGCCAGTGAGAGCCTCCGAAGGGGCAGGTA
CTGTGCCAGGGCAGCTCTGAAATTATGGATATTCTTATCCTCCTGGTTCCTTCGGTGCCA
ATGGTAACCTAATACCAGCCGCAGGAGCGCCATTTCTCCTAAAGGGCTACACCACTGTCA
ACATTATCCTGGACTCTGTGTCTCTCTGTGGGTCTTGTGGCATCACATCAGGCCAAA
ATTGCCAGACCAGGACCCTAAGTGTCTGATAGAGGCGATGATCTTTTCCAAAGTCAGTAC
30 TTA

 AA493341

GCTCGACTTTTTTTTTTGGGGGAACGTTTTTCATTAGGTAAACAGTGTTTGGCAAGCATTGG
AAACACGGAATCTCACAGACAGATACAGGCAGAAAGAATCACAGTTCAATCCAAAAGCAA
35 CACACTGAGAGGACATCAGAGTCCAAACACATGCAGAGAAGCTGTCAGGGAGCAGCTAGG
AGACACGCAGAGTTGCCTCACACGTGGCAGCAGGAGAAGGTGCAACACGGATCCGACTGC
TTACCCACTAAGGACACCAAGAACCAGGTTAAGGACGAAAAATGAGCCAAGGATGATCAG
ACTAACAAAATACACCCATGGCCATTCCCATCCTATCGCATCATTTACCCAGTAGAGCAC
GTCTGTCCAGCCCTCCATGGTGATGCACTGAAACACAGTAAGCATGGCAAAGGCAAAGTT
40 ATCAAAGTTGGTGATGCCTCCGTTCTGGGCCAACCCAGCCACTCCTACATTCCGTGCCATT

GGCAGTACACTGGCGTCCATTCCCTGT

AI467998

TTTTTTTTTTTTTTTTTGGTCCAAAATTTTAAATAGTATACAGACAACCTGTTAATTTTT
5 TTTTTTTTTTTTTTTGTAAATAACAAACACCACTTTGTTATGAAGACCTTACAAACCTCT
TCTTAAGACATTCTTACTCTGATCCAGGCAAAAACACTTCAAGGTTTGTAATGACTCTT
TCCTGACATAAATCCTTTTTTATTAAAATGCAAAATGTTCTTCAGAATAAACTGTGTAA
TAATTTTTTATACTTGGGAGTGCTCCTTGACACAGAGCTGTCATTTGCCAGTGAGAGCCTCC
GACGGGGCAGGTACTGTGCCAGGGCAGCTCTGAAATTATGGATATTCTTATCCTCCTGGT
10 TCCTTCGGTGCCAATGGTAACCTAATACCAGCCGCAGGGAGCGCCATTTCTCCTAAAGGG
CTACACCACTGTCAACATTATCC

AA885585

TTTTTTTTTTTTTTTTTCTTACAAAGAAAAATTTAATATTCGATGAGAGGTTGAACCAG
15 GCTTAAAGCAGACATACTAGGAAATGGTGCAGCCTGTAAGAATGCCAGTTTGTAAGTACT
GACTTTGGAAAAGATCATCGCCTCTATCAGACACTTAGGGTCCTGGTCTGGCAATTTTGG
CCTGATGTGATGCCACAAGACCCAACAGAGAGAGACACAGAGTCCAGGATAATGTTGACA
GTGGTGTAGCCCTTTAGGAGAAATGGCGCTCCCTGCGGCTGGTATTAGGTTACCATTGGC
ACCGA

20

AI033648

TGTAAATAACAAACACCACTTGGTTATGAAGACCTTACAAACCTCTTCTTAAGACATTCT
TACTCTGATCCAGGCAAAAACACTTCAAGGTTTGTAATGACTCTTTCCTGACATAAATC
CTTTTTTATTAAAATGCAAAATGTTCTTCAGAATAAACTGTGTAATAATTTTATACTT
25 GGGAGTGCTCCTTGACACAGAGCTGTCATTTGCCAGTGAGAGCCTCCGACGGGGCAGGTAC
TGTGCCAGGGCAGCTCTGAAATTATGGATATTCTTATCCTCCTGGTTCCTTCGGTGCCAA
TGGAACCTAATACCAGCCGCAGGGAGCGCCATTTCTCCTAAAGGGCTACACCACTGTCA
ACATTATCCTGGACTC

30 AI697633

ATTCCTGTTAATTTTGACAAGCTCAACGGCTGAAATCTAGGAATGGTTACTACCAAAGC
CCACCCAATCCAGCTCATTTTGCTATCGTTTTATAACAATTAATCTGCATTATATTTGGA
TCCAGACAAATAAAGCAATTATAAATGTATCTCACTTTACAACAGACAAAAAAGGGCAT
GCTATGGAAATTGTTTAAATCTCAAGCAACAATGCTGATTAATTTCTGGTCAATAATCGT
35 TCTATAGTTCTCCTTCATGAAGCCTGGTGAGGTTCCAGGAAACAGCTTGATTTGGGAAGC
CTCAGCAGAAAAGAAAGCATCTCAGAGGACACATAAAATGTCTGGCAACCCCTCTTGGCG
GCCCTCATCCAGCAAAGCTTGTGTGGTCTTGGCAACTGTCCTCAGGACTCTGCTTTCAAG
ATGAAAGAGGTGTAGCTTACCCGCTCAATACACCAAGTACAAGATTTAGTACGAAAAATG
ACCCAAAGATGACGAGACTGACACAATACACCCAGGGCAATTCAAATCCCATAGCATCAT
40 TCAT

AA523647

GGTCGACGTATTTGTAAAGAGATGGTCTATATCTTGTAATTACTGTATTGTATTTGAACT
GCAGCAATATCCATGGGTCCTAATAATTGTAGTTCCCCACTAAAATCTAGAAATTATTAG
5 TATTTTTACTCGGGCTATCCAGAAGTAGAAGAAATAGAGCCAATTCTCATTATTCAGCG
AAAATCCTCTGGGGTTAAAATTTTAAGTTTGAAAGAACTTGACACTACAGAAATTTTTCT
AAAATATTTTGAGTCACTATAAACCTATCATCTTCCACAAGAAAAAAAAACAAAAAA
AGTCGACG

BQ710377

CAAAGTACTTCCCCACATTTAGCTGGATTTGTCTTTGGTTTGAAGAGGCTAATACGTGAA
AGATTTGTTACAGTTGGATGTCCCCTTTTCTGAACCATGAAGTAATATTGTGAATGGAG
TTGAATGCTGAGGTTAGGGTGCCGGAAAGATTCAGGGTCCTTCGGTACCCTCACATGGCT
TGGCTTTGGTAGAACAAGAACTAAGCTCTGATTTGGCTTTAAATGAGAGTGCTAAATTT
15 CCTTTTTCTAATAAAGAACCTAGCTAAACATTTATATATACTTTTGAACACTGAACTTTC
TTGTTGCAGAGTTAACAGCTGTTGGGGGTAGCTGACAGCTGGATCCTGGTGCTGTTGGTA
CCATGGTACCTGAAGTGCACAGGCTGGTAGCCACACCTGACATTAACAAGTGAGTGGTAA
CCTCTCTGCCGCTGGCTCACAGCTACTGTTTCCATAGAAATGGCTGTCGGGATCAGTGGA
AACGAGGTAAGTGAAAGTTTTCGCTGATCCTTGTTTCCATCAAGCTGACGTCTGTTTCCC
20 TGGCAACAGCAGTGGACAGCAGCCAGGCGCTAGCAACAGATTCAGTAGAGCTCTCACTTG
TCAGCTGTGGCTATCATCTGTTCCCTGACCAAGTTCTTTTTTTTTTTTTTAATAATGTACA
GAAAGACCTCTGANGGACCAGGANGCNACTCTGGCCACATGTGCCCTCCTGGATGCTCGT
TTTGCAAATGGAGAGCTGTGTGCTGAGTTGACTTCTCTGTCCGCAGTTCCCCCTCCACTG
NGGCTCTGGGGTTGNTGATGTGCAGGTAAAAAAAGGAGGGTTGTTGAAGGTTATTAGTT
25 GTTCCAAGGGGAAGCCTGTTGAAACCTGGTTGATCCCCAATCCCTATGGGGAAGAAAAAT
CTCTTTAAGGGGCTTTTCATGCCCAGAGACCCAAATTTT

BQ706920

GGTGGCGATTTCGGACGAGGGCAAAGACTTCCCCCATTTAGCTGGATTTGTCTTTGGTTTG
30 AAGAGGCTAATACGTGAAAGATTTGTTACAGTTGGATGTCCCCTTTTCTGAACCATGAA
GTAATATTGTGAATGGAGTTGAATGCTGAGGTTAGGGTGCCGGAAAGATTCAGGGTCCTT
CGGTACCCTCACATGGCTTGGCTTTGGTAGAACAAGAACTAAGCTCTGATTTGGCTTTA
AATGAGAGTGCTAAATTTCTTTTTCTAATAAAGAACCTAGCTAAACATTTATATATACT
TTTGAACACTGAACTTTCTTGTTGCAGAGTTAACAGCTGTTGGGGGTAGCTGACAGCTGG
35 ATCCTGGTGCTGTTGGTACCATGGTACCTGAAGTGCACAGGCTGGTAGCCACACCTGACA
TTAACAAGTGAGTGGTAACCTCTCTGCCGCTGGCTCACAGCTACTGTTTCCATAGAAATG
GCTGTCGGGATCAGTGGAACGAGGTAAAGTGAAAGTTTTTCGCTGATCCTTGTTTCCATCA
AGCTGACGTCTGTTTCCCTGGCAACAGCAGTGGACAGCAGCCAGGCGCTAGCAACAGATT
CAGGAGAGCTCTCACTTGTCAGCTGTGGCTATCATCTGTTCCCTGACCAAGTTCTTTTTTT
40 TTTTTTTAATAATGGACAGAAAGACCTCTGAGGACCCAGGAGGCACCTCTGGGCACATGT
GCCCTCCTGGATGCTCCTTTTGCAGATGGAGACCTGGGGGCTGAGTTGACTTCTCTGGCC

GCAGTTCCCCCTCCACCTGGGGCTCCTGGGTGGTGAGGGGCCAGGTAAAAAAGGGAAGG
TGTTTGAGGGTATTAATGGGTCCCCGGGCGGGCTGATCGAATCCTGGGGACTCCACGTCC
CTGGGGGGACAAGAATCTCTTCAACGGGGTTTTCCGGCCGGGAGCCGGAGTTTTTTATTC
AGCGGG

5

BQ016847

TTTTTTTTTTTTTTTTTTTTCTTGTGGAAAGATGATAGGTTTATAGTGAATCAAAATATTTT
AGAAAAATTTCTGTAGTGTCAAGTTCCTTCAAACCTTAAAATTTTAACCCAGAGGATTTT
CGCTGAATAAATGAGAATTGGCTCTATTTCTTCTACTTCTGGATAGCCCGAGTAAAAATA
10 CTAATAATTTCTAGATTTTAGTGGGGAACCTACAATTATTAGGACCCATGGATATTGCTGC
AGTTCAAATACAATACAGTAATTACAAAATATAGACCATCTCTTTACAAATACAAATTAT
AGTATATTACAAGTCATGTACAGTAAATCTATAATTTTAAACAACTAGTGTATCTAAGT
TTACCTGGTTGCGAGTGCATTATTATTCCAGTTTACAGTTGCCCTTAGCGTGACAGTCAG
AAACCGACCATCGGAGTGATATTCTCTTATGTAAACTGGCGTCACATCACAGAAAACCTT
15 ATTTATGAGGTCCCATTTGCCCTCGCAATAATCACTGGTAGCTGGGTTCTGACTTACTTAC
ACACCGTATTTTCAAGACAGCTAAACAGGAACCAGGACGCAGTGTATTTGGGGGAAAGGGT
TTACAAATGGATATGTTGGGCCAGTGACTGATATGCTAGACCGATGGCTGAGGTAACGA
CACAGGTGTGATGATCGTCATCACCTTTAACT

20 **CA943595**

TGCAAATAAGGACAAGCTCAGCGGCTGAAATCTACAAATGGGGACTACCAAAGCCACC
CAATCCAGCTCATTTTGCTATCGTTTTATAACAATTAATCTGCATTATATTTGGATCCAG
ACAAATAAAGCAATTATAAATGTATCTCACTTTAGAACAGACAAAAAAGGGCATGCTAT
GGAAATTGTTTAAATCTCAAGCAACAATGCTGATTAATTTCTGGTCAATAATCGTTCTAT
25 AGTTCTCCTTCATGAAGCCTGGTGAGGTTCCAGGAAACAGCTTGATTTGGGAAGCCTCAG
CAGAAAAGAAAGCATCTCAGAGGACACATAAAATGTCTGGCAACCCCTCTTGGCGGCCCT
CATCCAGCAAAGCTTGTGTGGTCTTGGCAACTGTCTCAGGACTCTGCTTTCAAGATGAA
AGAGGTGTAGCTTACCCGCTCAATACACCAAGTACAAGATTTAGTACGAAAAATGACCCA
AAGATGACGAGACTGACAAAATACACCCAGGGCAATTCAAATCCCATAGCATCATTCATC
30 TGCAAGAAATAAGATGGTCTCATAGGAGTGGGTTAATAAGAGGATTTAATAAGGA

BM008196

GGCAAAGTACTTCCCCACATTTAGCTGGATTGGTCTTTGGTTTTGAAGAGGCTAATACGTG
AAAGATTTGTTTACAGTTGGATGTCCCCTTTTCTGAACCATGAAGTAATATTGTGAATGG
35 AGTTGAATGCTGACGGTTAGGGTGCCGGAAGATTCAAGGTCCTTCGGTACCCTCACATG
GCTTGGCTTTGGTAGAACAAGAACTAAGCTCTGATTTGGCTTTAAATGAGAGTGCTAAA
TTTCCTTTTTTCTAATAAAGAACCTAGCTAAACATTTATATATACTTTTGAACACTGAACT
TTCTTGTCTAGCAGAGTTAACAGCTGTAGGGGGTAGCTGACACGGCTGGATCCTGGTGTCTG
TTGGTACCATGGTACCTGAAGTGACAGGCTGGTAGCCACACCTGACATTAACAACGTGA
40 GTGGTAACCTCTCTGCCGCTGGCTCACAGCTACTGTTTCCATCAGAAATGGCTGTCTGGGC
TCACGTGGAAACGAGGTAAGTGAAAGTACGCTAGATCCTTGTTCCATCACAGCTGACGCT

CTGTTTCCCATGGCAACACCCAGCACGGACAAGCCGCCACGCCGCATAGACAACCACAAC
CACGTACAGCTCTCCACAAGTCAGCTCGTGGCTATCCATCATGTCCCTGAACAAGCCCAC
ACCACCCCCCCCCAAGCGACACAGCAACGAGCACCACCCGGACGAACCAAAGGACGGACC
CCCCTGCCCCAACCTCTCGCCCATCCGCGACAGACCCGCCAAGCAAACACGACAACCTAA
5 CAAAGCAGAGGGACAGACCCATAGCGCCCGCTACCGGAAGCGTACACCACTTCCCAACAG
TAAGGCCAAAAGAGCGACGCGGAGCACGTGAACGGATAAGAAAACGAGAGAAGGCACGGC
CGCATGGCAAACACACCAGCAAGCAGCAGACAGCACGTGGGCACGACACAGGACAGAAAG
CAGCCCACCTCAGAGGGGACCAACGAAGAGTCGCACGAC

10 BI769856

CTGGGCCCCAACATATCCATTTTTTAAACCCTTTCCCCCAAATACACTGCGTCCTGGTTCCCT
GTTTAGCTGTTCTGAAATACGGTGTGTAAGTAAGTCAGAACCCAGCTACCACTGATTATT
GCGAGGGCAATGGGACCTCATAAATAAGGTTTTCTGTGATGTGACGCCAGTTTACATAAG
AGAATATCACTCCGGTGGTCCGTTTCTGACTGTCACGCTAAGGGCAACTGTAACTGGAA
15 TAATAATGCACTCGCAACCAGGTAACTTAGATACTAGTTTGTTTTAAATTTATAGATT
TACTGTACATGACTTGTAATATACTATAATTTGTATTTGTAAAGAGATGGTCTATATTTT
GTAATTACTGTATTGTATTTGAACTGCAGCAATATCCATGGGTCCTAATAATTGTAGTTC
CCCACTAAAATCTAGAAATTATTAGTATTTTTTACTCGGGCTATCCAGAAGTAGAAGAAAT
AGAGCCAATTCTCATTTATTTCAGCGAAAATCCTCTGGGGTTAAAATTTTAAGTTTGAAAG
20 AACTTGACACTACAGAAATTTTTCTAAAATATTTTGAGTCACTATAAACCTATCATCTTT
CCACAAGATATAACCAGATGACTATTTGCAGTCTTTTCTTTGGGCAAGAGTTCCATGATTT
TGATACTGTACCTTTGGATCCACCATGGGTTGCAN

 BI758971

25 GGAAAAGAAATACTGTTTTAGAGAAATAACATTTTCAACAAAACATCCCTGGAGTCAGAT
TTTGAGTTGGGGTGGGCTAATCAGGGAGTCGGGGCTCTCTGCGTGATGTCAGTTCTATGG
CTAACTGGTTTTTTCTAAACCAGCCAGCTGCCTATCAAAACAGTACAACCTTTTCTAGGAAA
TGCAATTGGCAAAGACACTTACGATGCTGAGAAGTACACAAGGTGAACTGCTCCAGTTT
TTCTCATAGCAGGGTCAGCAGGAAAGCAAGTGGTGGCCCTGGTCCCATCTCACACAGGTG
30 AGACTGCACCGAGAGGTAACGTGGCCCTCACAGCCCACCACGCCTGGCCTTCGCCCAATT
CTGAAACTTCGTAGGATAGAGCTGGAAAGTGCCACATGGTGAAGCGAGATCCAGCTGTCT
GGGTGGATGTCGGAGTCCATAGGCTGAGCAGAGATGGTTCTTAGTGAGGTCTCGCTGCC
AGTTGACGGTGAAATCATAGCTGCCATTTACATTTTGTGAGATTATGAAAAACATAAGAC
TAAAGAAACTAAATGTGTTATTCCTGTGGACACAAAAATGTGTGTTTTTCAGATGGGGAG
35 GGGACCAAAAAGGAAAAACATTTTCATCTTAAACTTTTCCTAAGACAAAGGAAAAACAAAA
ACCATGCTCTACAACCTTCAAATTTTTCTTACAAAGAAAAATTTAATATTCGATGAGCAGG
TTGAACCAGGCTTAAAGCAGACATACTAGGAAATGGTGCAGCCTGTAAGAATGCCAGTTT
GTAAGTACTGACTTTGGAAAAGATCATCGCTCTATCAGACACTTAGGGTCCTGGTCTGGC
CATTTTGGCCTGATGTGATGCCAAAAGACC

40

 AA468565

TTTTATCGTGTGGAAAGATGATAGGTTTATAGTGACTCAAAATATTTTAGAAAAATTTCT
GTAGTGTCAAGTTCTTTCAAACCTAAAATTTTAACCCAGAGGATTTTCGCTGAATAAAT
GAGAATTGGCTCTATTTCTTCTACTTCTGGATAGCCCGAGTAAAAATACTAATAATTTCT
AGATTTTAGTGGGGAACCTACAATTATTAGGACCCATGGATATTGCTGCAGTTCAAATACA
5 ATACAGTAATTACAAAATATAGACCATCTCTTTACAAATACAAATTATAGTATATTACAA
GTCATGTACAGTAAATCTATTTTAAACAACTAGTGTATCTAAGTTTACCTGGTTGCGAG
TGCATTAT

AA437099

10 CTTACAAAGAAAAATTTAATATTCGATGAGAGGTTGAACCAGGCTTAAAGCAGACATACT
AGGAAATGGTGCAGCCTGTAAGAATGCCAGTTTGTAAGTACTGACTTTGGAAAAGATCAT
CGCCTCTATCAGACACTTAGGGTCTGGTCTGGCAATTTGGCCTGATGTGATGCCACAA
GACCCAACAGAGAGAGACACAGAGTCCAGGATAATGTTGACAGTGGTGTAGCCCTTTAGG
AGAAATGGCGCTCCCTGCGGCTGGTATTAGGTTACCATTGGCACCGAAGAGACCAGGAGG
15 ATAAGAATATCCATAATTTTACAGAGCTGCCCTGGCACAGTACCTGCCCCGTCGGAGGCTCT
CACTGGCAAATGACAGCTCTGTGCAAGGAGCACTCCCAAGTATAAAAATTAT

CA867864

20 CCGCGTCCGGTCAGATGGTACAAGTTTGTCTCTATAATTAAGACTTTTCCACCATCACAA
ACTTTAAACACAAAGTCTAAAATCTTGGGCAGCATAGAAAATAGGTTCTAGCTAAGCAGG
AGTTTTGTCTCTACCAAGACCTTTCCTGAAAATCACTTATCAAGACAGTTTCCTGTAA
AAAAAGCCATATCCAGCTGATTTTCCTTCCTGGGGCCAAAATCTGCTATTATTCGGCCT
GAAAGCCTTGATGACTCTGT
GTATGGATGCTTGTGTGTGTGTATGGGGAATATGTGATTAATGTGTGTTGGCTGCTGTTG
25 TCTCTGATTGGCTACTGTTGTTTCTGATTTAAATCTAAGTAAATGTTTAATTAAATGTA
TAGAATGCTGTCTCTAATGTGACCCTCTCTCCTTATTAAATCCTCTTATTAACCCACTCC
TATGAGACCATCTTATTTCTTGAGATGAATGATGCTATGGGATTTGAATTGCCCTGGGT
GTATTTTGTGAGTCTCGTCATCTTTGGGTCATTTTTCGTACTAAATCTTGTAAGTGGTGT
ATTGAGCGGG

30

AA682690

AATGCAAAATGTTCTTCAGAATAAACTGTGTAATAATTTTATACTTGGGATGTGCTCC
TTGCACAGAGCTGTCATTTGCCAGTGAGAGCCTCGACAGGCAGGTACTGTGCCAGGGCAG
CTCTGAAATTATGGATATTCTTATCCTCCTGGTTCCTTCTGTGCTCAATGGTAACCTAAT
35 ACCAGCCGCAGGACNCGCCATTTCTCCTAAAGGGCTACACCACTGTCAACATTATC

AA701888

TCAGCGAAAATCCTCTGGGGTTAAAATTTTAAGTTTGAAAGAACTTGACACTACAGAAAT
TTTTCTAAAATATTTTGAGTCACTATAAACCTATCATCTTCCACAAGATATAACAGATG
40 ACTATTTGCAGTCTTTCTTTGGGCAAGAGTTCATGATTTTGATACTGTACCTTGGAT

CCACCATGGGTTGCAACTGTCTTTGGTTTTGTTTGTGTTGACTTGAACCACCCTCTGGTAA
GTAAGTAAGTGAATTACAGAGCAGGTCCAGCTGGCTGCTCTGCCCCTTGGGTATCCATAG
TTACGGTTTTCTCTGTGGCCACCCAGGGTGTGTTTTTGCATCGCTGGTGCAGAAATGCAT
AGGTGGATGAGATATAGCTGCTCTTGTCTCTGGGGACTGGTGGTGTGCTTAAGAAATA
5 AGGGGTG

BU182632

TTTTGTCTAGTCTCGTCATCTTTGGGTCATTTTTTCGTAATAATCTTGTAAGTGGTGTATT
GAGCGGGCACAGTGGCTCACGCCTATAATCCCAGCACTTTCGGAGGCCGAGGCAGCTGGA
10 CCACCCGAGATCAGGAGTTTGAGACCAGCCTGACTAAGGCAGTGAAACCCTGTCTCTACT
AAAAATACAAAAATTAGCCAGGCATGGTGGCGCATGCCTGTAATCCCAGCTACTTGGGAG
GCTGAGGCAGGAGAATCACTTGAACCAGGGAGGTGGAGATTGCAGTGAGCCAAGACTGCA
CCATTGCATTCACGCCTGGGTGACAAGAGCAAACTCCATCTCAAAAAAAAAAAAAAAAAA
AAAAAAAAAAGACTTTTCTCTCATTCAACACTTTACCAGCATCTACTGACAGAAAATGG
15 ACAATTGAATTTCTCCAATATATATACCTCTGATATGTCTGCTTTGTAAAAGAGTAGTG
TAATTGCTTACAACATTGAAAAGGTTGTTATTGGGGTCTGGGGTAGCCAGGATATCGGC
ATGATTTGTCACCATATTCAGAATAAACTGTACTGCAATAGTGAGTTAATTCCATATCT
TGGCCAACAGAGAATTTTTGGCCAGTGGCTACTAAGGCACACGGAAGTCCAGTCTAAAAG
GGACAGGGGAGGACTCTTTGTAGATAGTTCTTATGATTAAAAATAACTTCCTATGTGTT
20 GTAGTGATGATTTAAGCTGACAGAATGCTAAAGACACCCCTTATGATTACCTGGTAGCAA
AGTACCTTCCCCACATTTAACCTGGATTTGCCCTTTTGGGTTTGAAAGAGGCTAAATA

BQ898429

GGTGGGATTCGGCACGAGGGCAAGACTTCCCCACATTTAGCTGGATTTGTCTTTGGTTTTG
25 AAGAGGCTAATACGTGAAAGATTTGTTACAGTTGGATGTCCCCTTTTCTGAACCATGAA
GTAATATTTGTGATATGGAGTTCGAATGGCTGAGGTCTAGGTGTGCCGAGAAAGATTGAG
GGTCCTTCGGTACCCTCACATGGCTTTGGCTTTGGTAGAACAAGAACTAAGCTCTGATTT
GGCTTTAAATGAGAGTGCTAAATTTCCCTTTTCTAATAAAGAACCTAGCTAAACATTTAT
ATATACTTTTGAACACTGAACTTTCTTGTGTCAGAGTTAACAGCTGTTGGGGGTAGCTGA
30 CAGCTGGATCCTGGTGCTGTTGGTACCATGGTACCTGAAGTGACAGGCTGGTAGCCACA
CCTGACATTAACAAGTGAGTGGTAACCTCTCTGCCGCTGGCTCACAGCTACTGTTTCCAT
AGAAATGGCTGTCGGGATCAGTGGAACGAGGTAAGTGAAAGTTTTGCTGATCCTTGTT
TCCATCAAGCTGACGTCTGTTTCCCTGGCAACAGCAGTGGACAGCAGCCAGGCGCTAGCA
ACAGATTCAGTAGAGCTCTCACTTGTGAGCTGTGGCTATCATCTGTTCCCTGACCAAGTTC
35 TTTTTTTTTTTTTTAATAATGTACAGAAAGACCTCTGAGGACCCAGGAGGCACCTCTGGC
CACATGTGCCCTCCTGGATGCTCGTTTTGCAGATGGAGAGCTGTGTGCTGAGTTGACTTC
TCTGTCCGCAGTTCCCCCTCCACCTGTGCTCTGGGTGTTGATGTGCCAGTTAAACAGG
GAGGCTGCTTCAGGGTATTAGTGTTGCCAAGGGGAGGCTGTTGAAATCTGTTGATCCCA
AATC
40

BQ711800

CAAAGTACTTCCCCACATTTAGCTGGATTTGTCTTTGGTTTGAAGAGGCTAATACGTGAA
AGATTTGTTACAGTTGGATGTCCCCTTTTCTGAACCATGAAGTAATATTGTGAATGGAG
TTGAATGCTGAGGTAGGGTGCCGGAAGATTCAGGGTCCTTCGGTACCCTCACATGGCT
TGGCTTTGGTAGAACAAGAACTAAGCTCTGATTTGGCTTTAAATGAGAGTGCTAAATTT
5 CCTTTTCTAATAAAGAACCTAGCTAAACATTTATATATACTTTTGAACACTGAACTTTC
TTGTTGCAGAGTTAACAGCTGTTGGGGGTAGCTGACAGCTGGATCCTGGTGCTGTTGGTA
CCATGGTACCCTGAAGTGACAGGCTGGTAGCCACACCTGACATTAACAAGTGAGTGGTAA
CCTCTCTGCCGCTGGCTCACAGCTACTGTTTCCATAGAAATGGCTGTCGGGATCAGTGGA
AACGAGGTAAAGTGAAGTTTTCGCTGATCCTTGTTTCCATCAAGCTGACGTCTGTTTCCC
10 TGGCAACAGCAGTGGACAGCAGCCAGGCGCTAGCAACAGATTCAGTAGAGCTCTCACTTG
TCAGCTGTGGCTATCATCTGTTCCCTGACCAAGTTCTTTTTTTTTTTTTTAATAATGTACA
GAAAGACCTCTGAGGACCCAGGGAGCACCTCTGGCCACATGTGCCCTCCTGAATGCTCGT
TTTGCAAATGGAGAGCTGTGTGCTGAGTTGACTTCTCTGTCCGCAGGTCCCCCTCCAAC
GTGCTCCTGGGTTGTGATGTGCAGGGTTAAACCAGGGAAGCTGTTGAAGGGTATTAGTGT
15 TGCCAGGGAAAGGCTGTTGAATTCTGGTTGATCCCAAATCCCTAGGGGGAAGAGAAATCC
CTTACGAGTGTTTTTTCATGGCCAGGAACCCTATA

AA703120

TCAGCGAAAATCCTCTGGGGTTAAAATTTTAAGTTTGAAAGAACTTGACACTACAGAAAT
20 TTTTCTAAAATATTTTGAGTCACTATAAACCTATCATCTTTCCACAAGATATAACCAGATG
ACTATTTGCAGTCTTTTCTTTGGGCAAGAGTTCCATGATTTTGATACTGTACCTTTGGAT
CCACCATGGGTGCAACTGTCTTTGGTTTTGTTTGTGTTGACTTGAACCACCCTCTGGTAA
GTAAGTAAGTGAATTACAGAGCAGGTCCAGCTGGCTGCTCTGCCCCTTGGGTATCCATAG
TTACGGTTTTTCTCTGTGGCCACCCAGGGTGTTTTTTGCATCGCTGGTGCAAAATGCAT
25 AGGTGGATGAGATATAGCTGCT

AA978315

GTATATCTTGTGGAAAGATGATAGGTTTATAGTGACTCAAAATATTTTAGAAAAATTTCT
GTAGTGTCAAGTCTTTTCAAACCTAAAATTTTAACCCAGAGGATTTTCGCTGAATAAAT
30 GAGAATTGGCTCTATTTCTTCTACTTCTGGATAGCCCGAGTAAAAATACTAATAATTTCT
AGATTTTAGTGGGGAACCTACAATTATTAGGACCCATGGATATAGCTGCAGTTCAAATACA
ATACAGTAATTACAAAATATAGACCATCTCTTTACAAATACAAATTATAGTATATTACAA
GTCATGTACAGTAAATCTATAATTTTAAACAACTAGTGTATCTAAGTTTACCAGGTTGC
GAGTGCATTATTATTCCAGTTTACAGTTGCCCTTAGCGTGACAGTCAGAAACCGACCATC
35 GGAGTGATATTCTCTTATGTAAACAGGCGTCACATCACAGA

BE550599

TTTTTTTTTTTGTGGAAAGATGATAGGTTTATAGTGACTCAAAATATTTTAGAAAAATTT
CTGTAGTGTCAAGTCTTTTCAAACCTAAAATTTTAACCCAGAGGATTTTCGCTGAATAA
40 ATGAGAATTGGCTCTATTTCTTCTACTTCTGGATAGCCCGAGTAAAAATACTAATAATTT
CTAGATTTTAGTGGGGAACCTACAATTATTAGGACCCATGGATATTGCTGCAGTTCAAATA

CAATACAGTAATTACAAAATATAGACCATCTCTTTACAAATACAAATTATAGTATATTAC
AAGTCATGTACAGTAAATCTATAATTTTAAACAAACTAGTGTATCTAAGTTTACCTGGTT
GCGAGTGCATTATTATTCCAGTTTACAGTTGCCCTTAGCGTGACAGTCAGAAACCGACCA
TCGGAGTGATATTCTCTTATGTAAACTGGCGTCACATCACAGAAAACCTTATTTATGAGG
5 TCCCATTGCCCTCGCAATAATCACTGGTAGCTGGGTTCTGACTTACTTACACACCGTATT
TCAGAACAGCTAAACAG

BE502741

TTTGGTATATCTTGTGGAAAGATGATAGGTTTATAGTGACTCAAAATATTTTAGAAAAAT
10 TTCTGTAGTGTCAAGTTCTTTCAAACCTTAAATTTTAAACCCAGAGGATTTTCGCTGAAT
AAATGAGAATTGGCTCTATTTCTTCTACTTCTGGATAGCCCGAGTAAAAATACTAATAAT
TTCTAGATTTTAGTGGGGAAC TACAATTATTAGGACCCATGGATATTGCTGCAGTTCAA
TACAATACAGTAATTACAAAATATAGACCATCTCTTTACAAATACAAATTATAGTATATT
ACAAGTCATGTACAGTAAATCTATAATTTTAAACAAACTAGTGTATCTAAGTTTACCTGG
15 TTGCGAGTGCATTATTATTCCAGTTTACAGTTGCCCTTAGCGTGACAGTCAGAAACCGAC
CATCGGAGTGATATTCTCTTATGTAAACTGGCGTCACATCACAGAAAACCTTATTTATGA
G

AW872382

TTTTTTTGTGGAAAGATGATAGGTTTATAGTGACTCAAAATATTTTAGAAAAATTTCTGT
20 AGTGTCAAGTTCTTTCAAACCTTAAATTTTAAACCCAGAGGATTTTCGCTGAATAAATGA
GAATTGGCTCTATTTCTTCTACTTCTGGATAGCCCGAGTAAAAATACTAATAATTTCTAG
ATTTTAGTGGGGAAC TACAATTATTAGGACCCATGGATATTGCTGCAGTTCAAATACAAT
ACAGTAATTACAAAATATAGACCATCTCTTTACAAATACAAATTATAGTATATTACAAGT
25 CATGTACAGTAAATCTATAATTTTAAACAAACTAGTGTATCTAAGTTTACCTGGTTGCGA
GTGCATTATTATTCCAGTTTACAGTTGCCCTTAGCGTGACAGTCAGAAACCGACCATCGG
AGTGATATTCTCTTATGTAAACTGGCGTCACATCACAGAAAACCTT

AW444663

CGGCCGCCAACTTTTTTGAATGAGTGAAGTGCCAGGTACCATGAGAAAACCCTAGCTGGT
30 AAAGATCAAACCTGAGTTAGTTCTAAATTCACATACGGATTTTTTTTGCATGACGAAATC
TATTCTCTTTTTCCTGACAACTTCTCCACCTAGATGTTTGGGAAAGTTGCCATGAGAGAT
AACAACCAGATCAATAGGAACAATAACTTCCAGACGTTTCCCCAGGCGGTGCTGCTGCTC
TTCAGGTGACTGCAACTGGCTTGGGCGGTGCTCCTGGGCAGGGGGGTCCGCTAGGCGTGG
35 GTCCAGAGGGACGGAGGACACAGGTTATTAAAGCAGTGTGCCTTTCTCAGTTG

AW341279

TAAATAACTAACACCATTTTGTATGAAGACCTTACAAACCTCTTCTTAAGACATTCTTA
CTCTGATCCAGGCAAAAACACTTCAAGGTTTGTAAATGACTCTTTCCTGACATAAATCCT
40 TTTTTTATTAAATGCAAAATGTTCTTCAGAATAAAACTGTGTAATAATTTTTTATACTTG

GGAGTGCTCCTTGACACAGAGCTGTCATTTGCCAGTGAGAGCCTCCGACGGGGCAGGTACT
GTGCCAGGGCAGCTCTGAAATTATGGATATTCTTATCCTCCTGGTTCCTTCGGTGCCAAAT
GGTAACCTAATACCAGCCGCAGGGAGCGCCATTTCTCCTAAAGGGCTACACCACTGTCAA
CATTATCCTGGACTCTGTGTCTCTCTCTGTTGGGTCTTGTGGCATCACATCAGGCCAAAA
5 TTGCCAGACCAGGACCCTAAGTGTCTGATAGAGGCGATGATCTTTTCCAAAGTCAGTACT
TACAAACTGGCATTCTTACAGGCTGCACCATTTCTAGTATGTCTG

CF456750

ACTTTTCTAGGAAATGCAATTGGCAAAGACACTTACGATGCTGAGAAGTACACAAGGTGA
10 AACTGCTCCAGTTTTTCTCATAGCAGGGTCAGCAGGAAAGCAAGTGGTGCCCCTGGTCCC
ATCTCACACAGGTGAGACTGCACCGAGAGGTAACGTGGCCCTCACAGCCCACCACGCCTG
GCCTTCGCCCAATTCTGAAACTTCGTAGGATAGAGCTGGAAAGTGCCACATGGTGAAGCG
AGATCCAGCTGTCTGGGTGGATGTCGGAGTCCATAGGCTGAGCAGAGATGGTTCTTAGTG
AGGTTCTCGCTGCCAGTTGACGGTGAAATCATAGCTGCCATTTACATTTTGTGAGATTAT
15 GAAAAACATAAGACTAAAGAACTAAATGTGTTATTCTGTGGACACAAAAATGTGTGTT
TTTCAGATGGGGAGGGGACCAAAAAGGAAAAACATTTTCATCTTAAACTTTCTTAAGACA
AAGGAAAAACAAAAACCATGCTCTACAACTTCAAATTTTTCTTACAAAGAAAAATTTAAT
ATTTCGATGAGAGGTTGAACCAGGCTTAAAGCAGACATACTAGGAAATGGTGCAGCCTGTA
AGAATGCCAGTTTGTAAAGTACTGACTTTGGAAAAGATCATCGCCTCTATCAGACACTTAG
20 GGTCTGGTCTGGCAATTTTGGCCTGATGTGATGCCACAAGACCCAACAGAGAGAGACAC
AGAGTCCAGGATAATGTTGACAGTGGTGTA

AW139850

TTTTTTTTTTTTTTTTTTTGAAGAAATAGAGCCAATTCTCATTTATTTCAGCGAAAATCCTC
25 TGGGGTTAAATTTTAAGTTTGAAAGAACTTGACACTACAGAAATTTTCTAAAATATTT
TGAGTCACTATAAACCTATCATCTTTCCACAAGATATAACCAGATGACTATTTGCAGTCTT
TTCTTTGGGCAAGAGTTCCATGATTTTGATACTGTACCTTTGGATCCACCATGGGTTGCA
ACTGTCTTTGGTTTTGTTTGTGTTGACTTGAACCACCCTCTGGTAAGTAAGTGAATTACAG
AGCAGGTCCAGCTGGCTGCTCTGCCCCTTGGGTATCCATAGTTACGGTTTTCTCTGTGGC
30 CCACCCAGGGTGTTTTTTGCATCGCTGGTGCAGAAATGCACAGGTGGATGAGATATAGCT
GCTCTTGTCTC

AW029633

TTATCTTGTGGAAAGATGATAGGTTTATAGTGACTCAAAATATTTTAGAAAAATTTCTGT
35 AGTGTCAAGTTCTTTCAAACCTTAAATTTTAACCCAGAGGATTTTCGCTGAATAAATGA
GAATTGGCTCTATTTCTTCTACTTCTGGATAGCCCGAGTAAAAATACTAATAATTTCTAG
ATTTTAGTGGGGAACACAAATTATTAGGACCCATGGATATTGCTGCAGTTCAAATACAAT
ACAGTAATTACAAAATATAGACCATCTCTTTACAAATACAAATTATAGTATATTACAAGT
CATGTACAGTAAATCTATAATTTTAAACAAACTAGTGTATCTAAGTTTACCTGGTTGCGA
40 GTGCATTATTATTCCAGTTTACAGTTGCCCTTAGCGTGACAGTCAGAAACCGACCATCGG
AGTGATATTCTCTTATGTAAACTGGCGTCACATCACAGAAAACCTTATTTATGAGGTCCC

ATTGCCCTCGCAATAATCACTG

AI963788

TTTTTCTTGTGGAAAGATGATAGGTTTATAGTGA CTCAAAATATTTTAGAAAAATTTCTG
5 TAGTGTCAAGTTCTTTCAAACCTTAAATTTTAAACCC CAGAGGATTTTCGCTGAATAAATG
AGAATTGGCTCTATTTCTTCTACTTCTGGATAGCCCGAGTAAAAATACTAATAATTTCTA
GATTTTAGTG GGGAACTACAATTATTAGGACCCATGGATATTGCTGCAGTTCAAATACAA
TACAGTAATTACAAAATATAGACCATCTCTTTACAAATACAAATTATAGTATATTACAAG
TCATGTACAGTAAATCTATAATTTTAAACAAACTAGTGTATCTAAGTTTACCTGGT

10

AI951788

ATCTTGTGGAAAGATGATAGGTTTATAGTGA CTCAAAATATTTTAGAAAAATTTCTGTAG
TGTCAAGTTCTTTCAAACCTTAAATTTTAAACCC CAGAGGATTTTCGCTGAATAAATGAGA
ATTGGCTCTATTTCTTCTACTTCTGGATAGCCCGAGTAAAAATACTAATAATTTCTAGAT
15 TTTAGTG GGGGAACCTACAATTATTAGGACCCATGGATATTGCTGCAGTTCAAATACAATA
CAGTAATTACAAAATATAGACCATCTCTTTACAAATACAAATTATAGTATATTACAAGTC
ATGTACAGTAAATCTATAATTTTAAACAAACTAGTGTATCTAAGTTTACCTGGTTGCGAG
TGCATTATTATTCCAGTTTACAGTTGCCCTTAGCGTGACAGTCAGAAACCGACCATCGGA
GTGATATTCTCTTATGTAAACT

20

AI680744

TTTTCTTCAAATAATTACAAGCTCAGCGGCTGAAATCTACAAATGGGGACTACCAAAGC
CCACCCAATCCAGCTCATTTTGCTATCGTTTTATAACAATTAATCTGCATTATATTTGGA
TCCAGACAAATAAAGCAATTATAAATGTATCTCACTTTAGAACAGACAAAAAAGGGCAT
25 GCTATGGAAATTGTTTAAATCTCAAGCAACAATGCTGATTAATTTCTGGTCAATAATCGT
TCTATAGTTCTCCTTCATGAAGCCTGGTGAGGTTCCAGGGAAACAGCTTGATTTGGGAAG
CCTCAGCAGAAAAGAAAGCATCTCAGAGGACACATAAAATGTCTGGCAACCCCTCTTGGC
GGCCCTCATCCAGCAAAGCTTGTGTGGTCTTG GCAACTGTCCTCAGGACTCTGCTTTCAA
GATGAAAGAGGTGTAGCTTACCCGCTCAATACACCAAGTACAAGATTTAGTACGAAAAAT
30 GACCCAAAGATGACGAGACTGACAAAATACACCCAGGGCAATTCAAATCCCATAGCATCA
TTCATCTGCAAG

AI601252

TTTGTAATAACAAACACCACTTTGTTATGAAGACCTTACAAACCTCTTCTTAAGACATT
35 CTTACTCTGATCCAGGCAAAAACACTTCAAGGTTTGTAATGACTCTTTCCTGACATAAA
TCCTTTTTTTATTAAATGCAAAATGTTCTTCAGAATAAACTGTGTAATAATTTTTATAC
TTGGGAGTGCTCCTTGACAGAGCTGTCATTTGCCAGTGAGAGCCTCCGACGGGGCAGGT
ACTGTGCCAGGGCAGCTCTGAAATTATGGATATTCTTATCCTCCTGGTTCTTCCGGTGCC
AATGGTAACCTAATACCAGCCGAGGGAGCGCCATTTCTCCTAAAGGGCTACACCACTGT
40 CAACATTATCCTGGACTCTGTGTCTCTCTCTGTTGGGTCTTGTGGCATCACATCAGGCCA

AAATTGCCAGACCAGGACCCTAAGTGTCTGATAGAGGCGATGATCTTTTCCAAAGTCAGT
ACTTACAAACT

AI459166

5 TTTTTTTTTTGGTCCAAAATTTTAAATAGTATACAGACAACCTGTTAATTTTTTTTTTTTTT
TTTTTTTTGTAAATAACAAACACCACTTTGTTATGAAGACCTTACAAACCTCTTCTTAAGA
CATTCTTACTCTGATCCAGGCAAAAACACTTCAAGGTTTGTAAATGACTCTTTCCTGACA
TAAATCCTTTTTTTATTAAAATGCAAAATGTTCTTCAGAATAAACTGTGTAATAATTTTT
ATACTTGGGAGTGCTCCTTGCACAGAGCTGTCATTTGCCAGTGAGAGCCTCCGACGGGGC
10 AGGTACTGTGCCAGGGCAGCTCTGAAATTATGGATATTCTTATCCTCCTGGTTCCTTCGG
TGCCAATGGTAACCTAATACCAGCCGCAGGGAGCGCCATT

AA885750

TCGACAGCTACCAGTGATTATTGCGAGGGCAATGGGACCTCATAAATAAGGTTTTCTGTG
15 ATGTGACGCCATTTACATAAGAGAATATCACTCCGATGGTCGGTTTCTGACTGTCACGCT
AAGGGCAACTGTAAACTGGAATAATAATGCACTCGCAACCAGGTAACTTAGATACACTA
GTTTGTTTAAATTATAGATTTACTGTACATGACTTGTAATATACTATAATTTGTATTTG
TAAAGAGATGGTCTATATTTTGTAATTACTGTATTGTATTTGAACTGCAGCAATATCCAT
GGGTCCTAATAATTGTAGTTCCCCACTAAAATCTAGAAATTATTAGTATTTTTACTCGGG
20 CTATCCAGAAGTAGAAGAAATAGAGCC

BX092736

GAATATGTGATTAATGTGTGTTGGCTGCTGTTGTCTCTGATTTGGCTACTGTTGTTTCTG
ATTTAAATCTAAGTAAATGTTTAAATTAATGTATAGAATGCTGTCTCTAATGTGACCCTC
25 TCTCCTTATTAAATCCTCTTATTAACCCACTCCTATGAGACCATCTTATTTCTTGAGAT
GAATGATGCTATGGGATTTGAATTGCCCTGGGTGTATTTTGTGAGTCTCGTCATCTTTGG
GTCATTTTTCTGACTAAATCTTGACTTGGTGTATTGAGCGGGTAAGCTACACCTCTTTC
ATCTTGAAAGCAGAGTCCTGAGGACAGTTGCCAAGACCACACAAGCTTTGCTGGATGAGG
GCCGCCAAGAGGGGTTGCCAGACATTTTATGTGTCTCTGAGATGCTTTCTTTTCTGCTG
30 AGGCTTCCCAAATCAAGCTGTTTCCTGGAACCTCACCAGGCTTCATGAAGGAGA

BX114568

TTTGTAATAACAAACACCACTTTGTTATGAAGACCTTACAAACCTCTTCTTAAGACATT
CTTACTCTGATCCAGGCAAAAACACTTCAAGGTTTGTAAATGACTCTTTCCTGACATAAA
35 TCCTTTTTTTATTAAAATGCAAAATGTTCTTCAGAATAAACTGTGTAATAATTTTTTATAC
TTGGGAGTGCTCCTTGCACAGAGCTGTCATTTGCCAGTGAGAGCCTCCGACGGGGCAGGT
ACTGTGCCAGGGCAGCTCTGAAATTATGGATATTCTTATCCTCCTGGTTCCTTCGGTGCC
AATGGTAACCTAATACCAGCCGCAGGGAGCGCCATTTCTCCTAAAGGGCTACACCACTGT
CAACATTATCCTGGACTCTGTGTCTCTCTGTTGGGTCTTGTGGCATCACATCAGGCCA
40 AAATTGCCAGACCAGGACCCTAAGTGTCTGATAGAGGCGATGATCTTTTCCAAAGTCAGT

ACTTACAAACTGGCATTCTTACAGGCTGCACCATTTCCTAGTATGTCTGCTTTAAGCCTG
GTTCAACCTCTCATCGAATATTAAATTTTCTTTGTAAGAAAAAAAAAAAAA

BE672659

5 TTTGTAAATAACAAACACCACTTTGTTATGAAGACCTTACAAACCTCTTCTTAAGACATT
CTTACTCTGATCCAGGCAAAAACACTTCAAGGTTTGTAATGACTCTTTCCTGACATAAA
TCCTTTTTTTATTAAATGCAAAATGTTCTTCAGAATAAACTGTGTAATAATTTTATAC
TTGGGAGTGCTCCTTGACAGAGCTGTCATTTGCCAGTGAGAGCCTCCGACAGGGCAGGT
ACTGTGCCAGGGCAGCTCTGAAATTATGGATATTCTTATCCTCCTGGTTCCTTCGGTGCC
10 AATGGTAACCTAATACCAGCCGCAGGGAGCGCCATTTCTCCTAAAGGGCTACACCACTGT
CAACATTATCCTGGACTCTGTGTCTCTCTCTGTTGGGTCTTGTGGCATCACATCAGGCCA
AAATTGCCAGACCAGGACCCTAAGTGTCTGATAGAGGCGATGATCTTTTCCAAAGTCAGT
ACTTACAAACTGGCATTCTTACAGGCTGCACCATTTCCTAGTATGTCTGCTTTAAGCCTG
GTTCAACC

15

N78509

GGAGAAAGGAGGGAAACCAGGAGCAGCCGGCATGGGCAGTGGCAGAATTGGCCCTGNTAG
AGAGCAGAGCTGATGCCATCCTTTTGGCAAATAGCTGACATTTTATGGTGTGGTGCTGGG
TGAGCCCCCTGTGAGGGTTGAACAGATGTGGACAGGACTTGGGTCCAGGCACTAGAGTGG
20 TGCAGCCTGTAAGAATGCCAGTTTGTAAGTACTGACTTTGGAAAAGATCATCGCCTCTAT
CAGACACTTAGGGTCTGGTCTGGCAATTTTGGCCTGATGTGATGCCACAAGACCCAACA
GAGAGAGACACAGAGTCCAGGATNAATGTTGACAGTGGTGTAGCCTTTAGGAAGAAATGG
CGCTCCCTGCGGCTGGTATTAGGTTACCATTGGCANCCGAAGGAACCCAGGAGGATTAAG
AATTTCCCTAATTTTCAAGACTTGCCCTGGCACAGTA

25

N73668

GGTCCAAAATTTTAAATAGTATACAGACAACCTGTAAATTTTTTTTTTTTTTTTTTTTGT
AAATAACAAACACCACTTTGTTATGAAGACCTTACAAACCTCTTCTTAAGACATTCTTAC
TCTGATCCAGGCAAAAACACTTCAAGGTTTGTAATCGACTCTTTCCTGACATAAATCCT
30 TTTTTATTAAAATNGCAAAATTGTTCTTCAGAATAAACTGTGTAATAATTTTATACTT
GGGAGTGCTCCTTGACAGAGCTGTCATTTGCCAGTGAGAGCCTCCGACGGGGCAGGTAC
TGTGCCAGGGCAGCTCTGAAATTATGGAAATTCCTATCCCCCTGGTTCCTNCGGTGGCCA
ATGGGTAACCTAATACCAGCCCGCGGAAGCGCCAATTTCNCCCAAAGGGGGTAAACCA
CTGGTNAAACATTA

35

N46744

TTTTTCTTTTGTAATAACAAACACCACTTTGTTATGAAGACCTTACAAACCTCTTCTTA
AGACATTCTTACTCTGATCCAGGCAAAAACACTTCAAGGTTTGTAATGACTCTTTCCTG
ACATAAATCCTTTTTTTATTAAATGCAAAATGTTCTTCAGAATAAACTGTGTAATAATT
40 TTTATANGTGGGGGNGCTC

N39597

ACAAAGAAAAATTTAATATTCGATGAGAGGTTGAACCAGGCTTAAAGCAGACATACTAGG
AAATGGTGCAGCCTGTAAGAATGCCAGTTTGTAAGTACTGACTTTGGAAAAGATCATCGC
5 CTCTATCAGACACTTAGGGTCTGGTCTGGCAATTTTGGCCTGATGTGATGCCACAAGAC
CCAACAGAGAGAGACACAGAGTCCAGGATAATGTTGACAGTGGTGTAGCCCTTTAGGAGA
AATGGCGCTCCCTGCGGCTGGTATTAGGTTACCATTTGGCACCGAAGAACCAGGAGGATAA
GAATATCCATAATTTTCAGAGCTTGCCCTGGCACAGTACCTGCCCCGTCGGAGGCTCTCAC
TGGGCAAATGGACAGCTCTGTGCAAGGAGCACTCCCAAGTATAANAATTATTACACAGTT
10 TTATTCTGAAGAACATTTTGCATTTTAAATAAAAAANGGA

BF439267

TTTTTTTTTTTTTTGGGCCAAAATTTTTAATAGTATACAGACAACCTGTTAATTTTTTTTT
TTTTTTTTTTTTTGTAAATAACAAACACCACTTTGTTATGAAGACCTTACAAACCTCTTTT
15 TAAGACATTCTTACTCTGATCCAGGCAAAAACACTTCAAGGTTTGTAAATGACTTTTTTCC
TGACATAAATCCTTTTTTTATTAATAATGCAAAATGTTCTTCAGAATAAAACTGTGTAATAA
TTTTTATACTTGGGAGTGCTCCTTGACACAGAGCTGTCATTTGCCAGTGAGAGCCTCCGAC
GGGGCAGGTACTGTGCCAGGGCAGCTCTGAAATTATGGATATTCTTATCCTCCTGGTTCC
TTCGGTGCCAATGGTAACCTAATACCAGCCGCAGGGAGCGCCATTTCTCCTAAAGGGCTA
20 CACCACTGTCAACATTATCCTGG

BF436153

TTTTTTTTTTGGTCCAAAATTTTTAATAGTATACAGACAACCTGTTAATTTTTTTTTTTTT
TTTTTTTTGTAAATAACAAACACCACTTTGTTATGAAGACCTTACAAACCTCTTCTTAAGA
25 CATTCTTACTCTGATCCAGGCAAAAACACTTCAAGGTTTGTAAATGACTCTTTCCTGACA
TAAATCCTTTTTTTATTAATAATGCAAAATGTTCTTCAGAATAAAACTGTGTAATAATTTTT
ATACTTGGGAGTGCTCCTTGACACAGAGCTGTCATTTGCCAGTGAGAGCCTCCGACGGGGC
AGGTACTGTGCCAGGGCAGCTCTGAAATTATGGATATTCTTATCCTCCTGGTTCTTCGG
TGCCAATGGTAACCTAATACCAGCCGCAGGGAGCGCCATTTCTCCTAAAGGGCTACACCA
30 CTGTCAACATTATCCTGGACTC

BF110611

TTTGTGGAAAGATGATAGGTTTATAGTGAATCAAAATATTTTAGAAAAATTTCTGTAGTG
TCAAGTTCTTTCAAACCTTAAATTTTAAACCCAGAGGATTTTCGCTGAATAAAATGAGAA
35 TTGGCTCTATTTCTTCTACTTCTGGATAGCCCGAGTAAAAATACTAATAATTTCTAGATT
TTAGTGGGGAACCTACAATTATTAGGACCCATGGATATTGCTGCAGTTCAAATACAATACA
GTAATTACAAAATATAGACCATCTCTTTACAAATACAAATTATAGTATATTACAAGTCAT
GTACAGTAAATCTATAATTTTAAACAACTAGTGTATCTAAGTTTACCTGGTTGCGAGTG
CATTATTATTCCAGTTTACAGTTGCCCTTAGCGTGACAGTCAGAAACCGACCATCGGAGT
40 GATATTCTCTTATGTAAACTGGCGTCACATCACAGAAAACCTTATTTATGA

M76558

```

5  gggcgagcgc ctccgtcccc ggatgtgagc tccggctgcc cgcggtcccg agccagcggc
   gcgcgggcgg cggcggcggg caccgggcac cgcggcgggc gggcagacgg gcgggcatgg
   ggggagcgcc gagcggcccc ggcgggccggg ccggcatcac cgcggcgtct ctccgctaga
   ggagggggaca agccagttct cttttgcagc aaaaaattac atgtatatat tattaagata
   atatatacat tggattttat ttttttaaaa agttttattt gctccatttt tgaaaaagag
   agagcttggg tggcgagcgg ttttttttta aaatcaatta tccttatttt ctgttatttg
   tccccgtccc tccccacccc cctgctgaag cgagaataag ggcagggacc gcggctccta
10 cctcttgggtg atcccccttc ccattccgcc cccgccccaa cggccagcac agtgccctgc
   acacagtagt cgctcaataa atgttcgtgg atgatgatga tgatgatgat gaaaaaatg
   cagcatcaac ggcagcagca agcggaccac gcgaacgagg caaactatgc aagaggcacc
   agacttcctc tttctgggtga aggaccaact tctcagccga atagctccaa gcaaactgtc
   ctgtcttggc aagctgcaat cgatgctgct agacaggcca aggctgccc aactatgagc
15 acctctgcac cccacctgt aggatctctc tcccaaagaa aacgtcagca atacgccaag
   agcaaaaaac agggtaactc gtccaacagc cgacctgcc gcgccctttt ctgtttatca
   ctcaataacc ccatccgaag agcctgcatt agtatagtgg aatggaaacc atttgacata
   tttatattat tggctatttt tgccaattgt gtggccttag ctatttacat cccattccct
   gaagatgatt ctaattcaac aaatcataac ttggaaaaag tagaatatgc cttcctgatt
20 atttttacag tcgagacatt tttgaagatt atagcgtatg gattattgct acatccta
   gcttatgtta ggaatggatg gaatttactg gattttgtta tagtaatagt aggattgttt
   agtgtaattt tggacaattt aaccaaagaa acagaaggcg ggaaccactc aagcggcaaa
   tctggaggct ttgatgtcaa agccctccgt gcctttcgag tgttgcgacc acttcgacta
   gtgtcaggag tgcccagttt acaagttgtc ctgaactcca ttataaaagc catggttccc
25 ctccctcaca tagccctttt ggtattatth gtaatcataa tctatgctat tataggattg
   gaacttttta ttggaaaaat gcacaaaaca tgtttttttg ctgactcaga tatcgtagct
   gaagaggacc cagctccatg tgcgttctca gggaatggac gccagtgtac tgccaatggc
   acggaatgta ggagtggctg ggttggcccc aacggaggca tcaccaactt tgataacttt
   gcctttgcca tgcttactgt gtttcagtgc atcaccatgg agggctggac agacgtgctc
30 tactggatga atgatgctat gggatttgaa ttgccctggg tgtattttgt cagtctcgtc
   atctttgggt catttttctg actaaatctt gtacttgggt tattgagcgg agaattctca
   aaggaaagag agaaggcaaa agcacgggga gatttcaga agctccggga gaagcagcag
   ctggaggagg atctaaaggg ctacttggat tggatcacc aagctgagga catcgatccg
   gagaatgagg aagaaggagg agaggaaggc aaacgaaata ctagcatgcc caccagcgag
35 actgagtctg tgaacacaga gaacgtcagc ggtgaaggcg agaaccgagg ctgctgtgga
   agtctctgtc aagccatctc aaaatccaaa ctcagccgac gctggcgtcg ctggaaccga
   ttcaatcgca gaagatgtag ggccgccgtg aagtctgtca cgttttactg gctggttatc
   gtccgtgtgt ttctgaacac cttaaccatt tcctctgagc actacaatca gccagattgg
   ttgacacaga ttcaagatat tgccaacaaa gtcctcttgg ctctgttcac ctgcgagatg
40 ctggtaaaaa tgtacagctt gggcctccaa gcatatttcg tctctctttt caaccggttt
   gattgcttcg tgggtgtgtg tggaaactc gagacgatct tgggtggaact ggaaatcatg
   tctcccctgg ggatctctgt gtttcgggtg gtgcgcctct taagaatctt caaagtgacc
   aggcactgga cttccctgtg caacttagtg gcatccttat taaactccat gaagtcagat
```

	gcttcgctgt	tgcttctgct	ttttctcttc	attatcatct	tttccttgct	tgggatgcag
	ctgtttggcg	gcaagtttaa	ttttgatgaa	acgcaaacca	agcggagcac	ctttgacaat
	ttccctcaag	cacttctcac	agtgttccag	atcctgacag	gcgaagactg	gaatgctgtg
	atgtacgatg	gcatcatggc	ttacgggggc	ccatcctctt	caggaatgat	cgtctgcatc
5	tacttcatca	tcctcttcat	ttgtggtaac	tatattctac	tgaatgtctt	cttggccatc
	gctgtagaca	atttggctga	tgctgaaagt	ctgaacactg	ctcagaaaga	agaagcggaa
	gaaaaggaga	ggaaaaagat	tgccagaaaa	gagagcctag	aaaataaaaa	gaacaacaaa
	ccagaagtca	accagatagc	caacagtgac	aacaaggtta	caattgatga	ctatagagaa
	gaggatgaag	acaaggaccc	ctatccgcct	tgcgatgtgc	cagtagggga	agaggaagag
10	gaagaggagg	aggatgaacc	tgaggttcct	gccggacccc	gtcctcgaag	gatctcggag
	ttgaacatga	aggaaaaaat	tgcccccctc	cctgaaggga	gcgctttctt	cattcttagc
	aagaccaacc	cgatccgcgt	aggctgccac	aagctcatca	accaccacat	cttcaccaac
	ctcatccttg	tcttcatcat	gctgagcagt	gctgccctgg	ccgcagagga	ccccatccgc
	agccactcct	tccggaacac	gatactgggt	tactttgact	atgccttcac	agccatcttt
15	actgttgaga	tcctgttgaa	gatgacaact	tttggagctt	tcctccacaa	aggggccttc
	tgcaggaact	acttcaattt	gctggatatg	ctgggtggtg	gggtgtctct	ggtgtcattt
	gggattcaat	ccagtgccat	ctccgttgtg	aagattctga	gggtcttaag	ggtcctgcgt
	cccctcaggg	ccatcaacag	agcaaaaagg	cttaagcacg	tgggtccagt	cgtcttcgtg
	gccatccgga	ccatcggcaa	catcatgatc	gtcaccaccc	tcctgcagtt	catgtttgcc
20	tgtatcgggg	tccagttggt	caaggggaag	ttctatcgct	gtacggatga	agccaaaagt
	aaccctgaag	aatgcagggg	acttttcatc	ctctacaagg	atggggatgt	tgacagtcct
	gtggtcctgt	aacggatctg	gcaaaacagt	gattttcaact	tcgacaacgt	cctctctgct
	atgatggcgc	tcttcacagt	ctccacgttt	gagggctggc	ctgcgttgct	gtataaagcc
	atcgactcga	atggagagaa	catcggccca	atctacaacc	accgcgtgga	gatctccatc
25	ttcttcatca	tctacatcat	cattgtagct	ttcttcatga	tgaacatctt	tgtgggcttt
	gtcatcgтта	catttcagga	acaaggagaa	aaagagtata	agaactgtga	gctggacaaa
	aatcagcgtc	agtgtgttga	atacgcttg	aaagcacgtc	ccttgcgagg	atacatcccc
	aaaaaccctt	accagtacaa	gttctgggtac	gtgggtgaact	cttcgccttt	cgaatacatg
	atgtttgtcc	tcatcatgct	caacacactc	tgcttggccca	tgcagcacta	cgagcagtc
30	aagatgttca	atgatgccat	ggacattctg	aacatggtct	tcaccggggg	gttcaccgtc
	gagatggttt	tgaagtcac	cgcatttaag	cctaaggggt	attttagtga	cgcctggaac
	acgtttgact	ccctcatcgt	aatcggcagc	attatagacg	tggccctcag	cgaagcagac
	ccaactgaaa	gtgaaaatgt	ccctgtccca	actgctacac	ctgggaactc	tgaagagagc
	aatagaatct	ccatcacctt	tttccgtctt	ttccgagtga	tgcgattggg	gaagcttctc
35	agcagggggg	aaggcatccg	gacattgctg	tggactttta	ttaagttctt	tcaggcgctc
	ccgtatgtgg	ccctcctcat	agccatgctg	ttcttcatct	atgcgggtcat	tggcatgcag
	atgtttggga	aagttgccat	gagagataac	aaccagatca	ataggaacaa	taacttccag
	acgtttcccc	aggcggtgct	gctgctcttc	aggtgtgcaa	caggtgaggc	ctggcaggag
	atcatgctgg	cctgtctccc	agggaaagctc	tgtgaccctg	agtcagatta	caacccccgg
40	gaggagcata	catgtgggag	caactttgcc	attgtctatt	tcatcagttt	ttacatgctc
	tgtgcatttc	tgatcatcaa	tctgtttgtg	gctgtcatca	tggataattt	cgactatctg
	accggggact	ggtctatttt	ggggcctcac	catttagatg	aattcaaaag	aatatggtca
	gaatatgacc	ctgaggcaaa	gggaaggata	aaacaccttg	atgtgggtcac	tctgcttcga
	cgcattccagc	ctcccctggg	gtttgggaag	ttatgtccac	acagggtagc	gtgcaagaga
45	ttagttgccca	tgaacatgcc	tctcaacagt	gacgggacag	tcatgtttta	tgcaaccctg

tttgcttttg ttcgaacggc tcttaagatc aagaccgaag ggaacctgga gcaagctaata
gaagaacttc gggctgtgat aaagaaaatt tggaagaaaa ccagcatgaa attacttgac
caagttgtcc ctccagctgg tgatgatgag gtaaccgtgg ggaagttcta tgccactttc
ctgatacagg actacttttag gaaattcaag aaacggaaag aacaaggact ggtgggaaag
5 taccctgcga agaacaccac aattgcccta caggcgggat taaggacact gcatgacatt
gggccagaaa tccggcgtgc tatatcgtgt gatttgcaag atgacgagcc tgaggaaaca
aaacgagaag aagaagatga tgtgttcaaa agaaatgggtg ccctgcttgg aaaccatgtc
aatcatgtta atagtgatag gagagattcc cttcagcaga ccaataccac ccaccgtccc
ctgcatgtcc aaaggccttc aattccacct gcaagtgata ctgagaaacc gctgtttcct
10 ccagcaggaa attcgggtgtg tcataaccat cataaccata attccatagg aaagcaagtt
cccacctcaa caaatgccaa tctcaataat gccaatatgt ccaaagctgc ccatggaaag
cgggccagca ttgggaacct tgagcatgtg tctgaaaatg ggcatcattc ttcccacaag
catgaccggg agcctcagag aaggtccagt gtgaaaagaa cccgctatta tgaaacttac
attaggtccg actcaggaga tgaacagctc ccaactatth gccgggaaga cccagagata
15 catggctatt tcagggaccc ccactgcttg ggggagcagg agtattttcag tagtgaggaa
tgctacgagg atgacagctc gccacctgg agcaggcaaa actatggcta ctacagcaga
taccagga gaaacatcga ctctgagagg ccccgaggct accatcatcc ccaaggattc
ttggaggacg atgactcgcc cgtttgctat gattcacgga gatctccaag gagacgccta
ctacctccca cccagcatc ccaccggaga tctccttca actttgagtg cctgcgccgg
20 cagagcagcc aggaagaggt cccgtcgtct cccatcttcc cccatcgcac ggccctgcct
ctgcatctaa tgcagcaaca gatcatggca gttgccggcc tagattcaag taaagcccag
aagtactcac cgagtcactc gaccgggtcg tgggccaccc ctccagcaac ccctccctac
cgggactgga caccgtgcta caccctctg atccaagtgg agcagtcaga ggccctggac
caggtgaacg gcagcctgcc gtccctgcac cgcagctcct ggtacacaga cgagcccagc
25 atctcctacc ggactttcac accagccagc ctgactgtcc ccagcagctt ccggaacaaa
aacagcgaca agcagaggag tgcggacagc ttggtggagg cagtcctgat atccgaaggc
ttgggacgct atgcaaggga cccaaaattt gtgtcagcaa caaaacacga aatcgctgat
gcctgtgacc tcaccatcga cgagatggag agtgcagcca gcaccctgct taatgggaac
gtgcgtcccc gagccaacgg ggatgtgggc cccctctcac accggcagga ctatgagcta
30 caggactttg gtccctggcta cagcgacgaa gagccagacc ctgggaggga tgaggaggac
ctggcgatg aaatgatatg catcaccacc ttgtagcccc cagcgagggg cagactggct
ctggcctcag gtggggcgca ggagagccag gggaaaagtg cctcatagtt aggaaagttt
aggcactagt tgggagtaat attcaattaa ttagactttt gtataagaga tgtcatgcct
caagaaagcc ataaacctgg taggaacagg tcccaagcgg ttgagcctgg cagagtacca
35 tgcgctcggc cccagctgca ggaaacagca ggccccgcc tctcacagag gatgggtgag
gaggccagac ctgccctgcc ccattgtcca gatgggcact gctgtggagt ctgcttctcc
catgtaccag ggcaccaggc ccaccaact gaaggcatgg cggcggggtg caggggaaag
ttaaaggtga tgacgatcat cacacctgtg tctttacctc agccatcggt ctagcatatc
agtactggg cccaacatat ccatttttaa accctttccc ccaaatacac tgcgtcctgg
40 ttctgtttta gctgttctga aatacgggtg gtaagtaagt cagaaccag ctaccagtga
ttattgcgag ggcaatggga cctcataaat aaggttttct gtgatgtgac gccagtttac
ataagagaat atcac

AF088004

	tttttttttt	cttacaaga	aaaattta	attcgatgag	aggttgaacc	aggcttaaag
	cagacatact	aggaaatggt	gcagcctgta	agaatgccag	tttgtaagta	ctgactttgg
	aaaagatcat	cgcctctatc	agacacttag	ggtcctggtc	tggcaatttt	ggcctgatgt
	gatgccacaa	gacccaacag	agagagacac	agagtccagg	ataatggtga	cagtgggtga
5	gcccttttagg	agaaatggcg	ctccctgcbg	ctggatttag	gttaccattg	gcaccgaagg
	aaccaggagg	ataagaatat	ccataatttc	agagctgccc	tggcacagta	cctgccccgt
	cggaggctct	cactggcaaa	tgacagctct	gtgcaaggag	cactcccaag	tataaaaatt
	attacacagt	tttattctga	agaacatttt	gcattttaat	aaaaaaggat	ttatgtcagg
	aaagagtcac	ttacaaacct	tgaagtgttt	ttgcctggat	cagagtaaga	atgtcttaag
10	aagaggtttg	taaggctctc	ataacaaagt	ggtgtttgtt	atttacaaaa	aaaaaaaaaa
	aaaaaaatta	acaggttgtc	tgtatactat	taaaaat		

M83566

	agaataagg	cagggaccgc	ggctcctatc	tcttggtgat	ccccttcccc	attccgcccc
15	cgcctcaacg	cccagcacag	tgccctgcac	acagtagtcg	ctcaataaat	gttcgtggat
	gatgatgatg	atgatgatga	aaaaaatgca	gcacacacgg	cagcagcaag	cggaccacgc
	gaacgaggca	aactatgcaa	gaggcaccag	acttcctctt	tctggtgaag	gaccaacttc
	tcagccgaat	agctccaagc	aaactgtcct	gtcttggcaa	gctgcaatcg	atgctgctag
	acaggccaag	gctgccccaa	ctatgagcac	ctctgcaccc	ccacctgtag	gatctctctc
20	ccaaagaaaa	cgctagcaat	acgccaagag	caaaaaacag	ggtaactcgt	ccaacagccg
	acctgccccg	gccctttttc	gtttatcact	caataacccc	atccgaagag	cctgcattag
	tatagtggaa	tggaaacccat	ttgacatatt	tatattattg	gctatttttg	ccaattgtgt
	ggccttagct	atttacatcc	cattccctga	agatgattct	aattcaacaa	atcataactt
	ggaaaaagta	gaatatgcct	tcctgattat	ttttacagtc	gagacatttt	tgaagattat
25	agcgtatgga	ttattgctac	atcctaatac	ttatgttagg	aatggatgga	atttactgga
	ttttgttata	gtaatagtag	gattgttttag	tgtaattttg	gaacaattaa	ccaaagaaac
	agaaggcggg	aaccactcaa	gcggcaaatc	tggaggcttt	gatgtcaaag	ccctccgtgc
	ctttcagagt	ttgcgaccac	ttcgactagt	gtcaggggtg	cccagtttac	aagttgtcct
	gaactccatt	ataaaagcca	tggttcccct	ccttcacata	gcccttttg	tattattttg
30	aatcataatc	tatgctatta	taggattgga	actttttatt	ggaaaaatgc	acaaaacatg
	tttttttgct	gactcagata	tcgtagctga	agaggaccca	gctccatgtg	cgttctcagg
	gaatggacgc	cagtgtactg	ccaatggcac	ggaatgtagg	agtggctggg	ttggcccga
	cggaggcatc	accaactttg	ataactttgc	ctttgccatg	cttactgtgt	ttcagtgcac
	caccatggag	ggctggacag	acgtgctcta	ctgggtaaat	gatgcgatag	gatgggaatg
35	gccatgggtg	tatttttgta	gtctgatcat	ccttggtcca	tttttcgtcc	ttaacctggt
	tcttggtgtc	cttagtgtag	aattctcaaa	ggaaagagag	aaggcaaaag	cacggggaga
	tttccagaag	ctccgggaga	agcagcagct	ggaggaggat	ctaaagggtc	acttggtattg
	gatcacccaa	gctgaggaca	tcgatccgga	gaatgaggaa	gaaggaggag	aggaaggcaa
	acgaaatact	agcatgccca	ccagcagagc	tgagtctgtg	aacacagaga	acgtcagcgg
40	tgaaggcgag	aaccgaggct	gctgtggaag	tctctggtgc	tggtggagac	ggagaggcgc
	ggccaaggcg	gggccctctg	ggtgtcggcg	gtgggtcaa	gccatctcaa	aatccaaact
	cagccgacgc	tggcgtcgct	ggaaccgatt	caatcgacga	agatgtaggg	ccgccgtgaa
	gtctgtcacg	ttttactggc	tggttatcgt	cctgggtgtt	ctgaacacct	taaccatttc

	ctctgagcac	tacaatcagc	cagattgggt	gacacagatt	caagatattg	ccaacaaagt
	cctcttggct	ctgttcacct	gcgagatgct	ggtaaaaaatg	tacagcttgg	gcctccaagc
	atatttcgct	tctcttttca	accggtttga	ttgcttcgctg	gtgtgtgggtg	gaatcactga
	gacgatcctg	gtggaactgg	aaatcatgtc	tcccctgggg	atctctgtgt	ttcgggtgtgt
5	gcgcctctta	agaatcttca	aagtgaccag	gcactggact	tccctgagca	acttagtggc
	atccttatta	aactccatga	agtcctatcg	ttcgtctgtg	cttctgcttt	ttctcttcat
	tatcatcttt	tccttgcttg	ggatgcagct	gtttggcggc	aagtttaatt	ttgatgaaac
	gcaaaccaag	cggagcacct	ttgacaattt	ccctcaagca	cttctcacag	tgttccagat
	cctgacaggc	gaagactgga	atgctgtgat	gtacgatggc	atcatggctt	acggggggccc
10	atcctcttca	ggaatgatcg	tctgcatcta	cttcatcatc	ctcttcattt	gtggtaacta
	tattctactg	aatgtcttct	tggccatcgc	tgtagacaat	ttggctgatg	ctgaaagtct
	gaacactgct	cagaaagaag	aagcgggaaga	aaaggagagg	aaaaagattg	ccagaaaaga
	gagcctagaa	aataaaaaaga	acaacaaacc	agaagtcaac	cagatagcca	acagtgacaa
	caaggttaca	attgatgact	atagagaaga	ggatgaagac	aaggaccctt	atccgccttg
15	cgatgtgcca	gtaggggaag	aggaagagga	agaggaggag	gatgaacctg	aggttctctgc
	cggaccccg	cctcgaagga	tctcggagtt	gaacatgaag	gaaaaaattg	cccccatccc
	tgaaggggagc	gctttcttca	ttcttagcaa	gaccaaccgc	atccgcgtag	gctgccacaa
	gctcatcaac	caccacatct	tcaccaacct	catccttgct	ttcatcatgc	tgagcagcgc
	tgccctggcc	gcagaggacc	ccatccgcag	ccactccttc	cggaacacga	tactgggtta
20	ctttgactat	gccttcacag	ccatctttac	tgttgagatc	ctggtgaaga	tgacaacttt
	tggagctttc	ctccacaaaag	gggccttctg	caggaactac	ttcaatttgc	tggatatgct
	ggtggttggg	gtgtctcttg	tgtcatttgg	gattcaatcc	agtgccatct	ccgttgtgaa
	gattctgagg	gtcttaaggg	tcctgcgtcc	cctcagggcc	atcaacagag	caaaaggact
	taagcacgtg	gtccagtgcg	tcttcgtggc	catccggacc	atcggcaaca	tcatgatcgt
25	cactaccctc	ctgcagttca	tgtttgcctg	tatcgggggtc	cagttgttca	aggggaagtt
	ctatcgctgt	acggatgaag	ccaaaagtaa	ccctgaagaa	tgcaggggac	ttttcatcct
	ctacaaggat	ggggatggtg	acagtcctgt	ggtccgtgaa	cggatctggc	aaaacagtga
	tttcaacttc	gacaacgtcc	tctctgctat	gatggcgctc	ttcacagtct	ccacgtttga
	gggctggcct	gcgttgctgt	ataaagccat	cgactcgaat	ggagagaaca	tcggcccaat
30	ctacaaccac	cgcgtggaga	tctccatctt	cttcatcatc	tacatcatca	ttgtagcttt
	cttcatgatg	aacatctttg	tgggctttgt	catcggttaca	tttcaggaac	aaggagaaaa
	agagtataag	aactgtgagc	tggacaaaaa	tcagcgtcag	tgtgttgaat	acgccttgaa
	agcacgtccc	ttgcggagat	acatccccaa	aaacccttac	cagtacaagt	tctggtacgt
	ggtgaactct	tcgccttttcg	aatacatgat	gtttgctctc	atcatgctca	acacactctg
35	cttggccatg	cagcactacg	agcagtccaa	gatgttcaat	gatgccatgg	acattctgaa
	catggtcttc	accgggggtg	tcaccgtcga	gatgggtttg	aaagtcatcg	catttaagcc
	taaggggtat	tttagtgacg	cctggaacac	gtttgactcc	ctcatcgtaa	tcggcagcat
	tatagacgtg	gccctcagcg	aagcggaccc	aactgaaagt	gaaaatgtcc	ctgtcccaac
	tgctacacct	gggaactctg	aagagagcaa	tagaatctcc	atcacctttt	tcctgtctttt
40	ccgagtgatg	cgattgggtga	agcttctcag	caggggggaa	ggcatccgga	cattgctgtg
	gacttttatt	aagtcctttc	aggcgctccc	gtatgtggcc	ctcctcatag	ccatgctgtt
	cttcatctat	gcggtcattg	gcatgcagat	gtttgggaaa	gttgccatga	gagataacaa
	ccagatcaat	aggaacaata	acttccagac	gtttccccag	gcggtgctgc	tgctcttcag
	gtgtgcaaca	ggtgaggcct	ggcaggagat	catgctggcc	tgtctcccag	ggaagctctg
45	tgaccctgag	tcagattaca	accccgggga	ggagtataca	tgtgggagca	actttgccat

	tgtctatttc	atcagttttt	acatgctctg	tgcattttctg	atcatcaatc	tgtttgtggc
	tgtcatcatg	gataatttcg	actatctgac	cggggactgg	tctatttttg	ggcctcacca
	tttagatgaa	ttcaaaagaa	tatggtcaga	atatgaccct	gaggcaaagg	gaaggataaa
	acaccttgat	gtggtcactc	tgcttcgacg	catccagcct	cccctggggg	ttgggaagtt
5	atgtccacac	agggtagcgt	gcaagagatt	agttgccatg	aacatgcctc	tcaacagtga
	cgggacagtc	atgtttaatg	caaccctggt	tgctttgggt	cgaacggctc	ttaagatcaa
	gaccgaaggg	aacctggagc	aagctaata	agaacttcgg	gctgtgataa	agaaaatttg
	gaagaaaacc	agcatgaaat	tacttgacca	agttgtccct	ccagctgggtg	atgatgaggt
	aaccgtgggg	aagttctatg	ccactttcct	gatacaggac	tacttttagga	aattcaagaa
10	acggaaagaa	caaggactgg	tgggaaagta	ccctgcgaag	aacaccacaa	ttgccctaca
	ggcgggatta	aggacactgc	atgacattgg	gccagaaatc	cggcgtgcta	tatcgtgtga
	tttgcaagat	gacgagcctg	aggaaacaaa	acgagaagaa	gaagatgatg	tgttcaaaag
	aatggtgcc	ctgcttgga	accatgtcaa	tcatgttaat	agtgatagga	gagattccct
	tcagcagacc	aataccaccc	accgtccctc	gcatgtccaa	aggccttcaa	ttccacctgc
15	aagtgatact	gagaaaccgc	tgtttcctcc	agcaggaaat	tcgggtgtgtc	ataaccatca
	taaccataat	tccataggaa	agcaagttcc	cacctcaaca	aatgccaatc	tcaataatgc
	caatatgtcc	aaagctgccc	atggaaagcg	gccagcatt	gggaaccttg	agcatgtgtc
	tgaaaatggg	catcattctt	cccacaagca	tgaccgggag	cctcagagaa	ggtccagtgt
	gaaaagaacc	cgctattatg	aaacttacat	taggtccgac	tcaggagatg	aacagctccc
20	aactattttgc	cgggaagacc	cagagataca	tggctatttc	agggaccccc	actgcttggg
	ggagcaggag	tatttcagta	gtgaggaatg	ctacgaggat	gacagctcgc	ccacctggag
	caggcaaaac	tatggctact	acagcagata	cccaggcaga	aacatcgact	ctgagaggcc
	ccgaggetac	catcatcccc	aaggattctt	ggaggacgat	gactcgcccc	tttgctatga
	ttcacggaga	tctccaagga	gacgcctact	acctcccacc	ccagcatccc	accggagatc
25	ctccttcaac	tttgagtgcc	tgcgcgggca	gagcagccag	gaagagggtcc	cgctcgtctcc
	catcttcccc	catcgcacgg	ccctgcctct	gcatctaata	cagcaacaga	tcatggcagt
	tgccggccta	gattcaagta	aagcccagaa	gtactcaccg	agtcactcga	cccggtcgtg
	ggccacccct	ccagcaaccc	ctccctaccg	ggactggaca	ccgtgctaca	ccccctgat
	ccaagtggag	cagtcagagg	ccctggacca	ggtgaacggc	agcctgccgt	ccctgcaccg
30	cagctcctgg	tacacagacg	agcccagacat	ctcctaccgg	actttcacac	cagccagcct
	gactgtcccc	agcagcttcc	ggaacaaaaa	cagcgacaag	cagaggagtg	cggacagctt
	ggtggaggca	gtcctgatat	ccgaaggctt	gggacgctat	gcaagggacc	caaaatttgt
	gtcagcaaca	aaacacgaaa	tcgctgatgc	ctgtgacctc	accatcgacg	agatggagag
	tgcagccagc	accctgctta	atgggaacgt	gcgtccccga	gccaacgggg	atgtgggccc
35	cctctcacac	cggcaggact	atgagctaca	ggactttggg	cctggctaca	gcgacgaaga
	gccagaccct	gggagggatg	aggaggacct	ggcggatgaa	atgatatgca	tcaccacctt
	gtagccccc	gcgaggggca	gactggctct	ggcctcaggt	ggggcgaggg	agagccaggg
	gaaaagtgcc	tcatagttag	gaaagttag	gcactagtgt	ggagtaatat	tcaattaatt
	agacttttgt	ataagagatg	tcatgcctca	agaaagccat	aaacctggta	ggaacagggtc
40	ccaagcgggt	gagcctggca	gagtaccatg	cgctcgcccc	cagctgcagg	aaacagcagg
	ccccgccctc	tcacagagga	tgggtgagga	ggccagacct	gccctgcccc	attgtccaga
	tgggactgc	tgtggagtct	gcttctccca	tgtaccaggg	caccaggccc	acccaactga
	aggcatggcg	gcgggggtgca	ggggaaagtt	aaaggtgatg	acgatcatca	cacctcgtgt
	cgttacctca	gccatcggtc	tagcatatga	gtcactgggc	ccaacatata	catttttaaa
45	ccctttcccc	caaatacact	gcgtcctggg	tcctgttttag	ctgttctgaa	ata

CB410657

GTACTGTGCCGGGGCAGCTCTGAAATTATGGATATTCTTATCCTCCTGGTTCCTTCGGTG
CCAATGGTAACCTAATAACCAGCCGCAGGGAGCGCCATTTCTCCTAAAGGGCTACACCACT
5 GTCAACATTATCCTGGACTCTGTGTCTCTCTCTGTTGGGTCTTGTGGCATCACATCAGGC
CAAAATTGCCAGACCAGGACCCTAAGTGTCTGATAGAGGCGATGATCTTTTCAAAGTCAG
TAC

BQ372430

10 TGCAGCAANTGGCACGGAATGTAGGAGTGGGTGGGTGGGACCGAACGGAGGCATCACCAA
CTTTGATAACTTGGCCTATGCCATGCTTACGGTGTTCAGTGCATCACCATGGAGGGCTG
GACAGATGTGCTCTACTGGGTAAATGATGCGATAGGATGGGAATGGCCATGGGCGTATTT
TGTTAGTCTGATCATCCTTGGCTCATTTTTTCGTCCTTAACCTGGTTCCTTGGTGTCTTAG
TGGAGAATTCTCAAAGGAAAGAGAGAAGGCAAAGCACGGGGAGATTTCCAGAAGCTCCG
15 GGAGAAGCAGCAGCTGGAGGAGGATCTAAAGGGCTACTTGG

BQ366601

ATGACTACGGGGGAAGTTCATTCTGACCTTCCAGACTAGCTAGTACTATATGAAATCCGA
GAGACGGAATGAACACGGACTGATGGGAAAGTACCCTGCGAAGAACACCACAATTGCCCT
20 ACAGGCGTGATTAAGGACACTGCATGATAGTTGCTCCAGAATGCCGGCGTGCTATATCGT
GTGATTTGCAAGATGACGAGCGTGAGGAAACAAAACGAGAAGAAGAAGATGATGTGTTCA
AAAGAAATGGTGGCCTGCTTGGAAACCATGTCAATCATGTTAATAGTGATAGGAGAGATT
CCCTTCAGCAGACCAATACCACCCACCGTCCNCTGCATGTCCAAAGGCCTTCAATTCCAC
CTGCAAGTGATACTGAGAAACCGCTGTTCTTCCAGCAGGAAATTCTG
25

BQ324528

TACATCTCCGCTATCTGTGCCGTGTAACACGGTGTCCAGTCTCGTTAGGGAGGGGCTGCT
GGAGGGGTGGCCACGACCGGGTCGAGTGACTCGGTGAGCACTTCTGTGCTTTACTTGAA
TCTAGGCCGGCAACTGCCATGATCTGTTGCTGCATTAGATGCAGAGGCAGTGCCGCGCGA
30 TGGTGAAGATGGGAGACGACGGGACCTCTTGCTGGCTGCTCTGCCGGCGCAGGCAC

BQ318830

TGTCGTGACTGGCGATACCTGGCGTTAGTGTGTACATGGTGTTTCATAATTGCTGCTGCAT
AACATTTTGTGAGAATTAATGTGACAATGTATGTGCAGTGCTTAGCACATAGCAAGTGCT
35 CATGAATGGTAGCCACCAAGATGGCTGTTGTCATTTTAGTTTGCAGCAGTTCCACTTGTC
ATCATTTAGTTCCCAGGGAGTCCCCTCTTCTTTGGGAACAGACTTGCTCTCTGTAGCTCC
ATTGCGGTAAAAACAGATGAGGTTAATCCCTGTCCCAATCATTTTGGAGATGGCGTCTGTT
TGTATTCCAATTCCACAGCCCAGTTCTTGTCTTTGTCTTCTTTTATTTAAGCAGCAGCC
ACACAGAATTAGCCCTTTTCAAAAATAAATAAGATTATCATCCTGTTTTGCGTCCCTGGG

GTAACAGACTCTAACATTTCTTTCTCTTTCTCTTTCTTTTCAGATTGTCTAGTGTAATTTTG
GAACAATTAACCAAAGAAACAGAAGGCGGGAACCACTCACGCGGCAAATCTGGAGGCTTT
GATGTCAAAGCCCTCCGTGCCTTTCGAGTGTTGCGACCACTTCGAA

5 AL708030

AGTTCCCACCTCAACAAATGCCAATCTCAATAATGCCAATATGTCCAAAGCTGCCCATGG
AAAGCGGCCCAGCATTGGGAACCTTGAGCATGTGTCTGAAAATGGGCATCATTCTTCCCA
CAAGCATGACCGGGAGCCTCAGAGAAGGTCCAGTGTGAAAAGGTCCGACTCAGGAGATGA
ACAGCTCCCAACTATTTGCCGGGAAGACCCAGAGATACATGGCTATTTTCAGGGACCCCCA
10 CTGCTTGGGGGAGCAGGAGTATTTTCAGTAGTGAGGAATGCTACGAGGATGACAGCTCGCC
CACCTGGAGCAGGCAAACTATGGCTACTACAGCAGATACCCAGGCAGAAACATCGACTC
TGAGAGGCCCCGAGGCTACCATCATCCCCAAGGATTCTTGAGGACGATGACTCGCCCGT
TTGCTATGATTACGGAGATCTCCAAGGAGACGCCTACTACCTCCCACCCCAGCATGTGA
GGCCAGATTTTTTTGTTTTTGGGTGGAACCTCCCGGGGAACAGTGTACCTTTCCCCCAACC
15 CCGCTCTG

BM509161

ATTCGGCACGAGCCTCCTTCAACTTTGAGTGCTCTGCCCCTTGGGTATCCATAGTTACGG
TTTTCTCTGTGGCCACCCAGGGTGTTTTTGCATCGCTGGTGCAGAAATGCACAGGTGG
20 ATGAGATATAGCTGCTCTTGCTCCTCTGGGGACTGGTGGTGCTGCTTAAGAAATAAGGGGT
GCTGGGGACAGAGGAGCAACGTGGTGATCTATAGGATTGGAGTGTGCGGGTCTGTACAAA
TCGTATTGTTGCCTTTTACAAAACCTGCTGTACTGTATGTTCTCTTTGAGGGCTTTTATAT
GCAATTGACTGAGGGCTGAAGTTTTTCATTAGAATGCACTCACACTCTGACTGTACGTCTT
GATGAAAACCCACTTTTGGATAATTAGAACCGTCAAGGCTTCATTTTCTGTCAACAGAAT
25 TAGGCCGACTGTCAGGTTACCTTGGCAGGGATTCCCTGCAATCAAAAAGATAGATGATAG
GTAGCAATTTTGGTCCAAAATTTTAAATAGTATACAGACAACCTGTTAATTTTTTTTTTT
TTTTTTTTTTGTAAATAACAAACACCACTTTGTTATGAAGACCTTACAAACCTCTT

N85902

30 GGAAAACCTCAAGTCCAGAGCAATACTACGTAAAATTCAGAAGTGAGAACATACAAAGGCA
ACACACAGGCTGACGAAGAAACAGAAAGAAGATACTGACCTGAGTTTGGATTTTGAGATG
GCTTGACTGAAAGAAAGACAAAAGTGTTAAGATTCTGGTTCCGAGGGCTTGAGCACACA
CTCCCCATCATTTTCAGCTGGAGATTTTCAT

35 BQ774355

TTTTTTTTTTTTTTTTTTTATTCTGAAGAACATTTTGCATTTTAATAAAAAAGGATTTATG
TCAGGAAAGAGTCATTTACAAACCTTGAAGTGTTTTTGCCTGGATCAGAGTAAGAATGTC
TTAAGAAGAGGTTTGTAAAGGTCTTCATAACAAAGTGGTGTTTGTTATTTACAAAAA
AAAAAAATTAACAGGTTGTCTGTATACTATTAAAAATTTTGGACCAAAATTGCTACCT
40 ATCATCTATCTTTTTGATTGCAGGGAATCCCTGCCAAGGTAACCTTGACAGTCGGCCTAAT

TCTGTTGACAGAAAATGAAGCCTTGACGGTCTAATTATCCAAAAGTGGGTTTTTCATCAG
GACGTACAGTCAGAGTGTGAGTGCATTCTAATGAAAACCTCTTCAGCCCTCATTCAATTG
CATACAAAAGCCCTCAAAGAGAACATACAGTACAGCAGTTTTGTAAAAGGCAACAATACG
ATTTGTACAGACCCCGACACTCCAATCCTATAGATCACCACGTTGCTCCTCTGTCCCCAG
5 CACCCCTTATTTCTTAAGCAGCACCACCAGTCCCCAGAGGACAAGAGCAGCTATATCTCA
TCCACCTGTGCATTTCTGCACCAGCGATGCANAAAACACCCTGGGGTGGGCCACAGAGAA
AACCGTAACTATGGATACCCAAGGGGC

CA774243

10 TAAATAACAAACACCCTTTGTTATGAAGACCTTACAAACCTCTTCTTAAGACATTCTTA
CTCTGATCCAGGCAAAAACACTTCAAGGTTTGTAAATGACTCTTTCCTGACATAAATCCT
TTTTTATTAAAATGCAAAATGTTCTTCAGAATAAACTGTGTAATAATTTTTTATACTTGG
GAGTGCTCCTTGACAGAGCTGTCATTTGCCAGTGAGAGCCTCCGACGGGGCAGGTACTG
TGCCAGGGCAGCTCTGAAATTATGGATATTCTTATCCTCCTGGTTCCTTCGGTGCCAATG
15 GTAACCTAATACCAGCCGAGGGAGCGCCATTTCTCCTAAAGGGCTACACCACTGTCAAC
ATTATCCTGGACTCTGTGTCTCTCTCTGTTGGGTCTTGTGGCATCACATCAGGCCAAAAT
TGCCAGACCAGGACCCTAAGTGTCTGATAGAGGCGATGATCTTTTCCAAAGTCAGTACTT
ACAAACTGGCATTCTTACAGGCTGCACCATTTCCTAGTATGTCTGCTTTAAGCCTGGTTC
AACCTCTCATCGAATATTAAATTTTTCTTTGTA

CA436347

TTTTTTTTTTTTTTTTCTTGGGGAAAGATGATAGGTTTATAGTGA CTCAAATATTTTAAA
AAAATTTCTGTAGGGTCAAGTTCTTTCAAACCTAAAATTTTAACCCAGAGGATTTTCGC
TGAATAAATGAAAATTGGCTCTATTTCTTCAACTTCGGGATAGCCCGAGTAAAAATACTA
25 ATAATTTCTAAATTTTAGGGGGGA ACTACAATTATTAGGACCCATGGATATTGCTGCAGT
TCAAATACAATACAGTAATTACAAAATATAGACCATCTCTTTACAAATACAAATTATAGT
ATATTACAAGTCATGTACAGTAAATCTATAATTTTAAACAAACTAGTGTATCTAAGTTTA
CCTGGTTGCGAGTGCATTATTATTCCAGTTTACAGTTGCCCTTAGCGTGACAGTCAGAAA
CCGACCATCGGAGTGATATTCTCTTATGTAAAC

CA389011

TGATTACTTGTAGCAAAGTACTTCCCCACATTTAGCTGGATTTGTCTTTGGTTTGAAGAG
GCTAATACGTGAAAGATTTGTTACAGTTGGATGTCCCCTTTTCTGAACCATGAAGTAAT
ATTGTGAATGGAGTTGAATGCTGAGGTTAGGGTGCCGAAAGATTCAGGGTCCTTCGGTA
35 CCCTCACATGGCTTGGCTTTGGTAGAACAAGAACTAAGCTCTGATTTGGCTTTAAATGA
GAGTGCTAAATTTCTTTTTCTAATAAAGAACCTAGCTAAACATTTATATATACTTTTGA
ACACTGAACNTCTTGTTGCAGAGTTAACAGCTGTTGGGGGTAGCTGACAGCTGGATCCT
GGTGCTGTTGGTACCATGGTACCTGAAGTGCACAGGCTGGTAGCCACACCTGACA

BU679327

TTTTTTTTTTTTTTTTCTTACAAAGAAAAATTTAATATTCGATNGAGAGGTTGAACCAGGC
TTAAAGCAGACATACTAGGAAATGGTGCAGCCTGTAAGAATGCCAGTTTGTAAAGTACTGA
CTTTGGAAAAGATCATCGCCTCTATCAGACACTTAGGGTCCTGGTCTGGCAATTTTGGCC
TGATGTGATGCCACAAGACCCAACAGAGAGAGACACAGAGTCCAGGATAATGTTGACAGT
5 GGTGTAGCCCTTTAGGAGAAATGGCGCTCCCTGCGGCTGGTATTAGGTTACCATTGGCAC
CGAAGGAACCAGGAGGATAAGAATATCCATAATTTTCAGAGCTGCCCTGGCACAGTACCTG
CCCCGTCTGGAGGCTCTCACTGGCAAATGACAGCTCTGTGCAAGGAGCACTCCCAAGTATA
AAAATTATTACACAGTTTTTATTCTGAAGAACATTTTGCATTTTAAATAAAAAAGGATTTAT
GTCAGGAAAGAGTCATTTACAAACCTTGAAGTGTTTTTGCCTGGATCAGAGTAAGAATGT
10 CTTAAGAAGAGGTTTGTAAAGTCTTCATAACANAGTGGTGTGTATTACAAAAAAA
AAAAAAAAAAAAATAAAAAAAAAAAAAAAAAAACCTCGTGCCGAATTCT

BU608029

TTTTTTTTTTTTTTTTGTAAATAACAAACACCACTTTGGTTATGAAGACCTTACAAACCT
15 CTTCTTAAGACATTCTTACTCTGATCCAGGCAAAAACACTTCAAGGTTTGTAAATGACTC
TTTCCTGACATAAATCCTTTTTTTATTTAAATGCAAAATGTTCTTCAGAATAAACTGTGT
AATAATTTTTATACTTGGGAGTGCTCCTTGACACAGAGCTGTCATTTGCCAGTGAGAGCCT
CCGACAGGGCAGGTACTGTGCCAGGGCAGCTCTGAAATTATGGATATTCTTATCCTCCTG
GTTCTTCGGTGCCAATGGTAACCTAATACCAGCCGAGGGAGCGCCATTTCTCCTAAAG
20 GGCTACACCACTGTCAACATTATCCTGGACTCTGTGTCTCTCTGTTGGGTCTTGTGGC
ATCACATCAGGCCAAAATTGCCAGACCAGGACCCTAAGTGTCTGATAGAGGCGATGATCT
TTTCCAAAGTCAGTACTTACAACTGGCATTCTTACAGGCTGCACCATTTCTAGTATGT
CTGCTTTAAGCCTGGTTCAACCTCTCATCGAATATTAATTTTTCTTTGTAAGAAAAATT
TGAAGTTGTAGAGCATGGTTTTTTGTTTTCCCTTGTCTTAGGAAAGTTTTAAGATGAAAT
25 GTTTTTCC

BU073743

AGTACACAAGGTGAAACTGCTCCAGTTTTTCTCATAGCAGGGTCAGCAGGAAAGCAAGTG
GTGCCCCTGGTCCCATCTCACACAGGTGAGACTGCACCGAGAGGTAACGTGGCCCTCACA
30 GCCCACCACGCCTGGCCTTCGCCCAATTCTGAACTTCGTAGGATAGAGCTGGAAAGTGC
CACATGGTGAAGCGAGATCCAGCTGTCTGGGTGGATGTCGGAGTCCATAGGCTGAGCAGA
GATGGTTCTTAGTGAGGTTCTCGCTGCCAGTTGACGGTGAAATCATAGCTGCCATTTACA
TTTTGTGAGATTATGAAAAACATAAGACTAAAGAACTAAATGTGTTATTCCTGTGGACA
CAAAAATGTGTGTTTTTCAGATGGGGAGGGGACCAAAAAGGAAAAACATTTTCATCTTAAA
35 ACTTCCCTAAGACAAAGGAAAACAAAAACCATGCTCTACAACCTCAAATTTTTCTTACA
AAGAAAAATTTAATAT

BE175413

AGCTGAGGAAACAAAACGAGAGAAGAAGATGATGTGTTCAAAGAAATGGTGCCCTGCTT
40 GGAAACCATGTCAATCATGTTAATAGTGATAGGAGAGATTCCCTTCAGCAGACCAATACC
ACCCACCGTCCCCTGCATGTCCAAAGGCCTTCAATTCCACCTGCAAGTGATACTGAGAAA

CCGCTGTTTCCTCCAGCAGGAAATTCGGTGTGTCATAACCATCATAACCATAATTCCATA
GGAAAGCAAGTTCCCACCTCAACAAATGCCAATCTCAATAATGCCAATATGTCCAAAGCT
GCCCATGGAAAGCGGCCAGCATAGGGAACCTTGAGCATGTGTCTGAAAATGGGCATCAT
TCTTCCCACAAGCATGACCGGGAGCCTCAGAGAAGGTCCAGTGTGAAAAGGTCCGACTCA
5 GGAGATGAACAGCTCCCAACTATTGGCCGGGAAGACCCAGAGATACATGGCTATTTTCAGG
CACCCCCACGGCTTGGGGGAGCAGGAGTATTTTCAGTAGTGAGGAATGCTACGAGGATGAC
AGCTCGCCACCTGGAGCAGGCAAACTATGGCTACTACAGCAGATACCCAGGCAGAAAC
ATCGACTCTGAGAGGCGCGAGGCTACATCATCCCAAGATTCTGGAGGAGATGACTCGCCG
TTTGTATGATCACGAGATCTCAAGAGAGCTATACTCCCACC

AW969248

TCTTGTGGAAAGATGATAGGTTTATAGTGACTCAAAATATTTTAGAAAAATTTCTGTAGG
GTCAAGTTCTTTCAAACCTTAAATTTTAAACCCAGAGGATTTTCGCTGAATAAATGAAAA
TTGGCTCTATTTCTTCTACTTCTGGATAGCCCGAGTAAAAATACTAATAATTTCTAGATT
15 TTAGTGGGGAAC TACAATTATTAGGACCCATGGATATTGCTGCAGTTCAAATACAATACA
GTAATTACAAAATATAGACCATCTCTTTACAAATCCAAATTATAGTATATTACAAGTCAT
GTACCGTAAATCTATTTTAAACAAACTAGGGTATCTAAGTTTACCTGGTTGCAAGTGCAT
TATTATTCCAGTTTACAGTTGCCCTTAGCGTGACAGTCAGAAACCGACCATCGGAGTGAT
ATTCTCTTATGTAAACTGGCGTCACATCACAGAAAACCTTATTTATTTGGGGGAAAGGGT
20 TTAAAAATGGATATGTTGGGCCCACTGACTGATAC

AI90811

GGAAAAGATCATCGCCTCTATCAGACACTTAGGGTCCTGGTCTGGCAATTTTGGCCTGAT
GTGATGCCACAAGACCCAACAGAGAGAGACACAGAGTCCAGGATAATGTTGACAGTGGTG
25 TAGCCCTTTAGGAGAAATGGCGCTCCCTGCGGCTGGTATTAGGTTACCATTTGGCACCGAA
GGAACCAGGAGGATAAGAATATCCATAATTTAGAGCTGCTCCTGGCACGGTACCTGCCCC
GTCCGAGGCTCTCACTGG

BF754485

GATGCGTGATGGCTGATCTAGAGGTATCCCATGGACTCTCATCGCAGCTCCTGGTACACA
GACGAGCCCGACATCTCCTACCGGACTTTACACCAGCCAGCCTGACTGTCCCCAGCAGC
TTCCGGAACAAAAACAGCGACAAGCAGAGGAGTGCGGACAGCTTGGTGGAGGCAGTCCTG
ATATCCGAAGGCTTGGGACGCTATGCAAGGGACCCAAAATTTGTGTCAGCAACAAAACAC
GAGATCGCTGATGCCTGTGACCTCACCATCGACGAGATGGAGAGTGCAGCCAGCACCCCTG
35 CTTAATGGGAACGTGCGTCCCCGAGCCAACGGGGATGTGGGCCCCCTCTCACACCGGCAG
GACTATGAGCTACAGGACTTTGGTCTTGGCTACAGCGACGAAGGGCCAGACCCTGGGAGG
GATGAGGAGGACCTGGCGGATGAAATGATATGCATCACACCTTGTAGCCCCCAGCGAGG
GGCAGACTGGCTCTGGCCTCAGGTGGGGCG

BI015409

CGCTCGTTTCGCTGTGCCAGGACAAAGTCCTGTAGCTCATAGTCCTGCCGTGTGAGAGGGG
GCCACATCCCCGTTNCTCGGGACGCACGACCCATTAAGCAGGGTGCTGGCTGCCCCCTCC
ATCTCGTCGATGGAGAGGTCANCAGGCATCAGCGATTTTCGTGTTTTGTGTGCGTGACACA
AATTTTGGGTCCCTTGCATACGCGTCCACAGCCTTACGGAGTATCAGCGACTGCTCTCC
5 ACCAATGCTGCCCCGCGACTCCTACTGCTTGTCGCTGTTTTTGGTTCCGGAAGCTGCTGG
GGACAGTCAGGCTGGCTGGTGTGAAAGTCCGGTAGGAGATGTCGGGCTCGTCTGTGTACC
AGGAGCTGCGGTGCAGGGACGGCAGGCTGCCGTTACCTGGTCCG

BG202552

10 GAGTTTCGAGCTTCTCTTTTCCTAAGNGAAAAAANAAGAANCACAAGNAAACCAAATAA
CCATGTTACTCTGTATAAAAATGCTAATCAGGGAATTCTGAATCAATAATGCTCCAATGA
AGGACAGAATTTAATTAGAAACAACACTAACCACAAGAGCCTAGCACAAACCCAACTCAG
AGCTTCCTGGTAATCTCAATGCGATGGATTCAATACAGACCATCTTATTTAAATTTCTC
15 ATCTGAGAGCTAATCAGCATTGAATGCATCATTTATTTTATGACACCAAAATTAAC TGCA
GTGATTCTTTAAGCATGGGGACACGTGACTCCCACTCTCAGCCCCGAGGGATGACAGCCA
AGAGCCTGGCTTCTGCCCAAGATTCCATCCGTTTTTGGTCTGCAGTGCATGGTCAACCATG
ATCCACAAAGCAGCAACCCGGGGGCTGTAGCTGCCGTGATGCGGGGGTAAGCCTGGCAGG
CTGCAACTGTTGCAGGGCTCCCAACACAGCCCCTGGACAAACGCGTCAGGGGAAAATAGG
GTTACCTGGCAATCTTTTTCTCTCTCTTTCTTCCGCTTCTTCTTTCTGAGCAGTGTTCA
20 GACTTTCAGCATCAGCCAAAGTGTCTACAGCGATGGCCAAGAAGACATTCAGTAGAATAT
CTAATTACAACTTTTTAAGGGCACAAACACACTACTAAATGCAACTACGTGCGGCCAACAA
TGGCAACGCCACACACCTCTGCATCCCGGAAGCTGGGTAGTAGGTGACGTCCCCAAGTG
TTATACTCACACAGCAAACCTAGAGTACCAGAGCCCTGCTTTTCAAACAANACANAACAA
ACAAACAACCCAAAGTAAAACCTGGTAAGGGACGTCTTCAGAAGTAAATTAC

BF883669

CTGGCTTTCCCATAGCACGCTCGGCAGGAAAGCAAGTGATGCCCCCTGGCTCCCATCTCAC
ACAGGTGACACTGCACCGAGAGGTAACGTGGCCCTCACAGCCCACCACGCCTGGCCTTCG
CCCAATTCTGAACTTCGTAGGATAGAGCTGGAAAGTGGCACATGGTGAAGCGAGATCCA
30 GCTGTCTGGGTGGATGTCCGAGCTCCATAGGCTGAGCAGAGATGGTTCTTAGTGAGGTTT
TCGCTGCCAGTTGACGGTGAAATCATAGCTGCCATTTACATTTTGTGAGATTATGAAAAA
CATAAGACTAAAGAACTAAATGTGTTATTCCTGTGGACACAAAAATGTGTGTTTTTCAG
ATGGGGAGGGGACCAAAAAGGAAAAACATTTTCATCTTAAACTTTCCTAAGACAAAGGAA
AACAA

BF817590

CTCAGCATGNATGAAACAGGATGAGGTTGGTGAAGATGTGGTGGTTGATGAGCTTGTGGC
AGCCTACGCGGATCGGGTTGGTCTTGCTAAGAATGAAGAAAGCGCTCCCTTCAGGGATGG
GGGCAATTTTTTCCTTCATGTTCAACTCCGAGATCCTTCGAGGACGGGGTCCGGCAGGAA
40 CCTCAGGTTTCATCCTCCTCCTCTTCTCTTCCCTACGGGCACATCGCAAGGCG
GATAGGGGTCCTTGTCTTCATCCTCTTCTCTATAGTCATCAATTGTAACCTTGTGTGCAC

TGTTGGCTATCTGGTTGACTTCTGGTTTGGTTGTTCTTTTTATTTTCTAGGCTCTCTTTTC
TGGCAATCTTTTTTCTCTCCTTTTTCTTCCGCTTCTTCTTTCTGAGCAGTGTTTCAGACTTT
CAGCATCAGCCAAATGGTCTA

5 **BF807128**

TCAAAGTCGAAGGAGGATCTCCGCGTGGGATGCTGGGGTGGGAGGTAGTAGGCGTCTCCT
TGGAGATCTCCGTGAATCATAGCAAACGGGCGAGTCATCGTCCTACAAGAATCCTAGTGG
ATGATGGTAGCCTCGGGGCCTCTCAGAGTCGATGTTTCTGCCTGG

10 **BF806160**

CTCGCCCGTTTGCTATGAGTCACGGAGATCTCCAAGGAGACGCCTACTACCTCCCACCCC
AGCATCCCACCGGAGATCCTCCTTCAACTTTGAGTGCCTGCGCCGGCAGAGCAGCCAGGA
AGAGGTCCCGTTCGTCTCCCATCTTCCCCCATCGCACGGCCCTGCCTCTGCATCTAATGCA
GCAACAGATCATGGCAGTTGCCGGCCTAGATTCAAGTAAAGCCCAGAAGTACTCACCGAG
15 TCACTCGACCCGGCCGTGGGCCACCCCTCCAGCAACCCCTCCCTACCGGGACTGGACACC
GTGCTACACCCCCCAGATGACGCCGATGTA

BF805244

CCAGGCAGAAACATCGACTCTGAGAGGCCCCGAGGCTACCATCATCCCCAAGGATTCTTG
20 GAGGACGATGACTCGCCCGTTTGCTATGATTCACGGAGATCTCCAAGGAGACGCCTACTA
CCTCCCACCCCAGCATCCCACCGGAGATCCTCCTTCAACTTTGAGTGCCTGCGCCGGCAG
AGCAGCCAGGAAGAGGTCCCGTCGTCTCCCATCTTCCCCCATCGCACGGCCCTGCCTCTG
CATCTAATGCAGCAACAGATCATGGCAGTTGCCGGCCTAGATTCAAGTAAAGCCCAGAAG
TACTCACCGAGTCACTCGACCCGGTCGTGGGCCACCCCTCCAGCAACCCCTCCCTACCGG
25 GACTGGACACCGTGCTACACCCCCCAGATGACGCCGATGTA

BF805235

TACATCGGCGTCATCTGGGGGGTGTAGCACGGTGTCCAGTCCCGGTAGGGAGGGGTTGCT
GGAGGGGTGGCCACGACCGGGTCGAGTGACTCGGTGAGTACTTCTGGGCTTTACTTGAA
30 TCTAGGCCGGCAACTGCCATGATCTGTTGCTGCATTAGATGCAGAGGCAGGGCCGTGCGA
TGGGGGAAGATGGGAGACGACGGGACCTCTTCCTGGCTGCTCTGCCGGCGCAGGCACTCA
AAGTTGAAGGAGGATCTCCGTTGGGATGCTGGGGTGGGAGGTAGTAGGCGTCTCCTTGGA
GATCTCCGTGAATCATAGCANACGGGCGAGTCATCGTCCTCCAAGAATCCTTGNNGATGA
TGGTAGCCTCGGNGCCTCTCAGAGTCGATGTTTCTGCCTGNGTATCTGCTCGGGCGAGCC
35 GGTACCGAGCT

BF805080

TACATCGGCGTCATCTGGGGGGTGTAGCACGGTGTCCAGTCCCGGTAGGGAGGGGTTGCT
GGAGGGGTGGCCACGACCGGGTCGAGTGACTCGGTGAGTACTTCTGGGCTTTACTTGAA

TCTAGGCCGGCAACTGCCATGATCTGTTGCTGCATTAGATGCAGAGGCAGGGCCGTGCGA
TGGGGGAAGATGGGAGACGACGGGACCTCTTCCTGGCTGCTCTGCCGGCGCAGGCACTCA
AAGTTGAAGGAGGATCTCCGGTGGGATGCTGGGGTGGGAGGTAGTAGGCGTCTCCTTGA
GATCTCCGTGAATCATAGCAAACGGGCGAG

5

T27949

GCGGACAGCTTGGTGGAGGCAGTCCTGATATCCGAAGCCTTNGGACGCTATGCAAGGGAC
CCAAAATTTNTTTCAGCAACAAAACACGAAATCGCTGATGCCTGTAACCTCACCATCGAC
GAGATGGAGAGTNCAGCCAGCACCCCTGCTTAATGGGAACGTGCGTCCCCGAGCCAACGGG
GAT

10

BE836638

AAGAAATAGGAGGATAAGAATATCATATTTTCAGAGCTGCCCTGGCACAGTACCTGCCCCG
TCGGAGGCTCTCACTGGCAAATGACAGCTCTGTGCAAGGAGCACTCCCAAGTATAAAAAT
TATTACACAGTT

15

BE770685

CCATTGGTACGAGAGAAATTAGGAGGATAAGATTATCTATTATTCTGAGCTGCCCTGGCA
CAGTACCTGCCCCGTCGGAGGCTCTCACTGGCAAATGACAGCTCTGTGCAAGGAGCACTC
CCAAGTATAAAAATTATTACATAGTTTTTATTCTGAAGAACATTTTGCATTTTAATAAAAA
AGGATTTATGTCAGGAAAGAGTCATTTACATACCTTGAATTGTTTTTGCCTGGATCAGAG
TAAGAATGTCTTAAGAAGAGGTTTGTAAAGTCTTCATAACAAAGTGGTGTGTTGTTATTTA
CAAAAAAAAAAAAAAAAAAAAAATTTTTATACCGGGTTTGTCTGTATACAAATTTCTCTG

20

BE769065

TCCAGAGTAGAAGAAATCAGCCAAGTATCATTTATTCAGCGAAAATCCTCTGGGGATTAA
AATTTTAAGTTTGAAAGAACTTGACACTACAGAAATTTTCTAAAATATTTTGAGTCACT
ATAAACCTATCATCTTTCCACAAGATATACCAGATGACTATTTGCAGTCTTTTCTTTGGG
CAAGAGTTCCATGATTTTGATACTGTACCTTTGGATCCACCATGGGTTGCAACTGTCTTT
GGTTTGTGTTGTTTGAACCAACCTCTGGAAAGCTACTCTGGAAA

25

30

Sequences identified as those of HOXB13 cluster

BF676461

GGGATTCCCCCGGCTGGGTGGGGAGAGCGAGCTGGGTGCCCCCATAGATTCCCCTGCCCC
AACCTCATGAGCCGACCCTCGGCTCCATGGAGCCCGGAAATTATGCCACCTTGGATGGAG
CCAAGGATATCGAAGGCTTGTTGGGAGCGGGAGGGGGCGGAATCTGGTCGCCCACTCCC
TCTCTGACCAGCCACCCAGCGCGCTACGCTTGATGCCTGTGTCAATATGCCCCCTTGATC

35

TGCCAGGCTCGGGGAGCGGCCAAAAGCAATGCCACCCCTATGCTCTGGGGGTGCCAGGG
GACTGTCCCCGGCTCCGTGCCTTATGGTTACTGTGGGGCGGGTACATACTCCTGCAGAG
TTGTCCCGGAGCTCGTTGAAACCTTGTGCCGAGGAGAGCCACCCTGGCGGTACCCGGGAA
GACTCCCCAGGGCGGGGAAGAGTACCCAGCGGCCCAATGAGTTGTGCTTCTATCGGGATA
5 TCCGGGACCTACCAGGCCTATGTGCAGGTACTGGACGTGTCCTGTGCTGCAGACTCTGGG
TGTCCGTGGAGACCGGACATTGGCTCGCTGTGGCCTGTGGCCGGTACCAGTCTTGGGCT
CTCGGTGTGTGGCTGGACACGCCGGTTGTGTTTCGCGGGAGACCGCACCCACCAGGTTCTT
TTGGGAGGGCCGCTTTGCAGACTCCGGGGGAGGCCCCCTCTGAGGCGGGGCCCTTTTCGGGG
GGGCGAAGAAAGCTTTCCGACGCAGGCGCTTGCGGAGCTGGCGGGACATCGGGACACTTC
10 ACCCAGCGAAGCGCGGCTTGGGGCCCCCTCTGGGCGCGGTCTCGGTTGACACCGGCGAAGA
GTTTCGGGAGAGGCCCATATCTTCTGGGGAGGGCGTTGCGTCGCCCCCG

BC007092

ggattcccc ggccctgggtg gggagagcga gctgggtgcc ccctagattc cccgcccccg
15 cacctcatga gccgaccctc ggctccatgg agcccggcaa ttatgccacc ttggatggag
ccaaggatat cgaaggcttg ctgggagcgg gaggggggcg gaatctggtc gccactccc
ctctgaccag ccaccagcg ggcctacgc tgatgcctgc tgtcaactat gccccttgg
atctgccagg ctcggcggag ccgccaaagc aatgccacc atgccctggg gtgccccagg
ggacgtcccc agctcccgtg ccttatggtt actttggagg cgggtactac tctgcccag
20 tgtcccggag ctcgctgaaa ccctgtgccc aggcagccac cctggcccg taccgcccg
agactccac ggccggggaa gaggaccca gccgccccac tgagtttgcc ttctatccg
gatatccggg aacctaccag cctatggcca gttacctgga cgtgtctgtg gtgcagactc
tgggtgctcc tggagaaccg cgacatgact ccctgttgcc tgtggacagt taccagtctt
gggctctcgc tgggtggctgg aacagccaga tgtgttgcca gggagaacag aaccaccag
25 gtcccttttg gaaggcagca tttgcagact ccagcgggca gcaccctcct gacgcctgcg
cctttcgtcg cggccgcaag aaacgcattc cgtacagcaa ggggcagttg cgggagctgg
agcgggagta tgcggctaac aagttcatca ccaaggacaa gaggcgcaag atctcggcag
ccaccagcct ctcgagcgc cagattacca tctggtttca gaaccgccg gtcaaagaga
agaaggttct cgccaagggt aagaacagcg ctaccctta agagatctcc ttgcctgggt
30 gggaggagcg aaagtggggg tgcctgggg agaccaggaa cctgccaagc ccaggctggg
gccaaggact ctgctgagag gcccctagag acaacaccct tcccaggcca ctggctgctg
gactgttcct caggagcggc ctgggtacc agtatgtgca gggagacgga acccatgtg
acagcccact ccaccagggt tcccaaagaa cctggcccag tcataatcat tcctcctgac
agtggcaata atcacgataa ccagtactag ctgccatgat cgttagcctc atattttcta
35 tctagagctc tgtagagcac tttagaaacc gctttcatga attgagctaa ttatgaataa
atttggaataa aaaaaaaaaa aaaaaaaaaa aaaaaa

BM462617

ATTCCCCGGCCTGGGTGGGGAGAGCGAGCTGGGTGCCCCCTAGATTCCCCGCCCCCGCA
40 CCTCATGAGCCGACCTCGGCTCCATGGAGCCCGCAATTATGCCACCTTGGATGGAGCC
AAGGATATCGAAGGCTTGCTGGGAGCGGGAGGGGGGCGGAATCTGGTCGCCCCACTCCCCT
CTGACCAGCCACCCAGCGGCGCCTACGCTGATGCCTGCTGTCAACTATGCCCCCTTGGAT

CTGCCAGGCTCGGCGGAGCCGCCAAAGCAATGCCACCCATGCCCTGGGGTGCCCCAGGGG
ACGTCCCCAGCTCCCGTGCCTTATGGTTACTTTGGAGGCGGGTACTACTCCTGCCGAGTG
TCCCGGAGCTCGCTGAAACCCCTGTGCCAGGCAGCCACCCTGGCCGCGTACCCCGCGGAG
ACTCCACAGGCCCGGGGAAGAGTACCCAGCCGCCCACTGAGTTTGCCTTCTATCCGGGA
5 TATCCGGGAACCTACCAGCCTATGGCCAGTTACCTGGACGTGTCTGTGGTGACAGACTCTG
GGTGCTCCTGGAGAACC GCGACATGACTCCCTGTTGCCTGTGGACAGTTACCAGTCCTGG
GCTCTCGCTGGTGGCTGGAACAGCCAGATGTGTTGCCAGGGAGAACAGAACCCACCAGGT
CCCCTTTTGGAAGGCAGCATTTGCAGACTCCAGCGGGCAGCACCCCTCCTGACGCCTGCGC
CTTTCGT

BG752489

GCAGGCGACTTGCGAGCTGGGAGCGATTTAAAACGCTTTGGATTCCCCCGGCCTGGGTGG
GGAGAGCGAGCTGGGTGCCCCCTAGATTCCCCGCCCCCGCACCTCATGAGCCGACCCTCG
GCTCCATGGAGCCCCGCAATTATGCCACCTTGGATGGAGCCAAGGATATCGAAGGCTTGC
15 TGGGAGCGGGAGGGGGGCGGAATCTGGTCGCCCACTCCCTCTGACCAGCCACCCAGCGG
CGCTACGCTGATGCCTGTGTCAACTATCCCCCTTGGATCTGCCAGGCTCGGCGGAGC
CGCCAAAGCAATGCCACCCATGCCCTGGGGTGCCCCAGGGGACGTCCCCAGCTCCCGTGC
CTTATGGTTACTTTGGAGGCGGGTACTACTCCTGCCGAGTGTCCCGGAGCTCGCTGAAAC
CCTGTGCCCAGGCAGCCACCCTGGCCGCGTACCCCGCGGAGACTCCCACGGCCGGGGAAG
20 AGTACCCCGAGCCGCCCACTGAGTTTGCCTTCTATCCGGGATATCCGGGAACCTACCAGC
CTATGGCCAGTTACCTGGACGTGTCTGTGGTGCAGACTCTGGGTGCTCCTGGAGAACCGC
GACATGACTCCCTGTTGCCTGTGGACAGTTACCAGTCTTGGGCTCTCGCTGGTGGCTGGA
ACAGCCAGATGTGTTGCCAGGGAGAACAGAACCCAGGTCCCTTTTTGGAAGGCAGCAT
CTGCAGACTCCAGCGGGCAGGACCTCCTGACGCCTGCGGCCTTTCGTGCGGAGCGCAAGA
25 AACGCATTCCGTA

BG778198

GGATTTAAAACGCTTTGGATTCCCCCGGCCTGGGTGGGGAGAGCGAGCTGGGTGCCCCCT
AGATTCCCCGCCCCCGCACCTCATGAGCCGACCCTCGGCTCCATGGAGCCCGCAATTAT
30 GCCACCTTGGATGGAGCCAAGGATATCGAAGGCTTGTGGGAGCGGGAGGGGGGCGGAAT
CTGGTCGCCCCACTCCCTCTGACCAGCCACCCAGCGGCGCCTACGCTGATGCCTGTGTG
AACTATGCCCCCTTGGATCTGCCAGGCTCGGCGGAGCCGCCAAAGCAATGCCACCCATGC
CCTGGGGTGCCCCAGGGACGTCCCCAGCTCCCGTGCCTTATGGTTACTTTGGAGGCGGGT
ACTACTCCTGCCGAGTGTCCCGGAGCTCGCTGAAACCCCTGTGCCAGGCAGCCACCCTGG
35 CCGCGTACCCCGCGGAGACTCCACGGCCGGGGAAGAGTACCCAGCCGCCCACTGAGT
TTGCCTTCTATCCGGGATATCCGGGAACCTACCAGCCTATGGCCAGTTACCTGGACGTGT
CTGTGGTGCAGACTCTGGGTGCTCCTGGAGAACC GCGACATGACTCCCTGTTGCCTGTGG
ACAGTTACCAGTCTTGGGCTCTCGCTGGTGGGCTGGAACAGCCAGATGTGTTGCCAGCGC
AGAACAGAACCCACCAGGTCCCTTTTTGGAAGGCAGCATTTGCAGACTCCAGCGGGCAGAA
40 CCCTCCTGACGCCTGCGCCTTTCGTTGCGGGCGAAAAA

CB050884

AAGAAACGCATTCCGTACAGCAAGGGGCAGTTGCGGGAGCTGGAGCGGGAGTATGCGGCT
AACAAAGTTCATCACCAAGGACAAGAGGCGCAAGATCTCGGCAGCCACCAGCCTCTCGGAG
CGCCAGATTACCATCTGGTTTCAGAACCGCCGGGTCAAAGAGAAGAAGGTTCTCGCCAAG
GTGAAGAACAGCGCTACCCCTTAAGAGATCTCCTTGCCTGGGTGGGAGGAGCGAAAAGTGG
5 GGGTGTCTCTGGGGAGACCAGGAACCTGCCAAGCCCAGGCTGGGGCCAAGGACTCTGCTGA
GAGGCCCTAGAGACAACACCCTTCCCAGGCCACTGGCTGCTGGACTGTTCTCAGGAGC
GGCCTGGGTACCCAGTATGTGCAGGGAGACGGAACCCCATGTGACAGCCCACTCCACCAG
GGTTCCCAAAGAACCTGGCCCAGTCATAATCATTTCATCCTGACAGTGGCAATAATCACGA
TAACCAGTACTAGCTGCCATGATCGTTAGCCTCATATTTTCTATCTAGAGCTCTGTAGAG
10 CACTTTAGAAACCGCTTTCATGAATTGAGCTAATTATGAATAAATTTGGAAGGCGAAAAA
AAAAACCTCGTGCC

CB050885

ATTCGGCACGAGGTTTTTTTTTTTCGCCTTCCAAATTTATTCATAATTAGCTCAATTCATG
15 AAAGCGGTTTCTAAAGTGCTCTACAGAGCTCTAGATAGAAAATATGAGGCTAACGATCAT
GGCAGCTAGTACTGGTTATCGTGATTATTGCCACTGTCAGGATGAATGATTATGACTGGG
CCAGGTTCTTTGGGAACCCCTGGTGGAGTGGGCTGTCACATGGGGTTCCGTCTCCCTGCAC
ATACTGGGTACCCAGGCCGCTCCTGAGGAACAGTCCAGCAACCAGTGGCCTGGGAAGGGT
GTTGTCTCTAGGGGCCTC

20

BF965191

GGGTGGGGAGAGCGAGCTGGGTGCCCCCTAGATTCCCCGCCCCCGCACCTCATGAGCCGA
CCCTCGGCTCCATGGAGCCCGGCAATTATGCCACCTTGGATGGAGCCAAGGATATCGAAG
GCTTGCTGGGAGCGGGAGGGGGGCGGAATCTGGTGCCTTCCCTCTGACCAGCCACC
25 CAGCGGCGCCTACGCTGATGCCTGCTGTCAACTATGCCCCCTTGGATCTGCCAGGCTCGG
CGGAGCCGCCAAAGCAATGCCACCCATGCCCTGGGGTGGCCAGGGGACGTCCCCAGCTC
CCGTGCCTTATGGTTACTTTGGAGGCGGGTACTACTCCTGCCGAGTGTCCCGGAGCTCGC
TGAAACCCTGTGCCAGGCAGCCACCCTGGCCGCGTAACCCGACGGAGACTCTCACGTGCG
GGGAAGAGTACCCCTAGCGCCCCACATGAGTTTGCCTTCTATCCGGGATATCCGGGACCG
30 TACCAGCCTATGGCAGTTACCTGGACGTGTCTGTGGTGGCGACTCTGGGTGCTCCTGGAG
AACC GCGGACATGACTCCTTGTTTGCTGTGCGACGCTCACCAGTCTGGGCTCCTCGTCCG
TGGTCGCACTCCCACTTTTGGCCGGGCGACATCCCCGGGGCCCCCTTCCGGAACAGCGAC
CTTGCGAGCCCCCGGGGACACACCCCGTAAGCGGCCTATCATCGCTGATAAACCTCATC
AGAGGGCACCGAAAGCCGCGACTCTAACCCCCCACTACGACTCACGACCGCACAGGTAC
35 TCGAACCGCCCAATATCTGGTTCTAACCCATGGCGCATCTCAGCCGCTAGAGAGCCAACC
AAACGCGCCACGCGCAACCACACTACACCACGGCACCCCTTTCATCTCACTCCCACGCCG
ATCACTCTTACCCCTCCAGAATCATTTCCCTCGCACATCCTACCTATCTCATGCCTCCCA
GTTACCCCCATTCCCTCCCCCTAATCTCACCCACACATTACGACGTTTCTCACTACGCTT
CGCTCCGACCCACATCCTCACCCCAACATTACATACCACTTACCATCACGACCCCCCTT
40 CTCATCGACTCCTGTCTCATTTCTCAACCACAGTACTACCAGCTCCAACACACCCTCACC
CCAAGCTATCCATCACCTACAGCTTTTACCCCTCACCGCTCCCAAGTAATTCAGATCAC
TCAAACACAATCTGCTACATACTCATCCCTCCCCCACTCCAGTACAGTCCAACCACCGA

CCA ACTACCTCCGCGCCACCCGCGCCGCCCCACCTCACCGGCCCCAACCGCCCGCACAGG
GCACGCACCCCCCGGCAACCGCGCGATCCGGCCGTACACACTCTTGGGCGGCACGCAGCT
GAGGACATTCCGCGGGAGCGCCCCACCGTGGGCTACGTGGGTGCGGACCCGGCGGGGCGC
GTGCGGCGTCGCCCCGCCCGCCGCGACTGCGACCCAGTCGAG

5

BU930208

GGGGCTTTGGATTCCCCCGGCCTGGGTGGGGAGAGCGAGCTGGGTGCCCCCTAGATTCCC
CGCCCCCGCACCTCATGAGCCGACCCTCGGCTCCATGGAGCCCGGCAATTATGCCACCTT
GGATGGAGCCAAGGATATCGAAGGCTTGCTGGGAGCGGGAGGGGGCGGAATCTGGTCGC
10 CCACTCCCCCTTGACCAGCCACCCAGCGGCGCCTACGCTGATGCCTGCTGTCAACTATGC
CCCCCTGGATCTGCCAGGCTCGGCGGAGCCGCCAAAGCAATGCCACCCATGCCCTGGGGT
GCCCCAGGGGACGTCCCCAGCTCCCGTGCCCTTATGGTTACTTTGGAGGCGGGTACTACTC
CTGCCGAGTGTCGCCGAGCTCGCTGAAACCCCTGTGCCAGGCAGCCACCCCTGGCCGCGTA
CCCCGCGGAGACTCCCACGGCCGGGGAAGAGTACCCAGCCGCCCCACTGAGTTTGCCTT
15 CTATCCGGGATATCCGGGAACCTACCAGCCTATGGCCAGTTACCTGGACGTGTCTGTGGT
GCAGACTCTGGGTGCTCCTGNAGAACCGCGACATGACTCCCTGTTGCCTGTGGACAGTTA
CCAGTCTTGGGCTCTCGCTGGTGGCCTGGAACAGCCCAGATGTGTTTGCCAGGGNAGAA
CACGAACCCCAACCCGTTCCCCCTTTTGGGAAAGGGCAGCCATTTTGGCCAGCCTTCCAA
GCGGGGCCAACCACCCCTCCCCTGGACAGGCCCTGGT

20

AA807966

GCGGCCGCAAGAAACGCATTCCGTACAGCAAGGGGCAGTTGCGGGACTGGAGCGGGAGTA
TGCGGCTAACAAGTTCATCACCAAGGACAAGAGGCGCAAGATCTCGGCAGCCACCAGCCT
CTCGGAGCGCCAGATTACCATCTGGTTTCAGAACCGCCGGGTCAAAGAGAAGAAGTTCT
25 CGCCAAGGTGAAGAACAGCGCTACCCCTTAAGAGATCTCCTTGCCCTGGGTGGGAGGAGCG
AAAGTGGGGGTGTCTGGGGAGACCAGGAACCTGCCAAGCCCAGGCTGGGGCCAAGGACT
CTGCTGAGAGGCCCTAGAGACAACACCCTTCCCAGGCCACTGGCTGCTGGACTGTTCTT
CAGGAGCGGCCTGGGTACCCAGTATGTGCAGGGAGACGGAACCCCATGTGACAGCCCATT
CCACCAGGGTTCCCAAAGAACCTGGCCCAGTCATAATCATTCATCCTGACAGTGGC

30

AI884491

AGCGGCCGCAAGAAACGCATTCCGTACAGCAAGGGGCAGTTGCGGGAGCTGGAGCGGGAG
TATGCGGCTAACAAGTTCATCACCAAGGACAAGAGGCGCAAGATCTCGGCAGCCACCAGC
CTCTCGGAGCGCCAGATTACCATCTGGTTTCAGAACCGCCGGGTCAAAGAGAAGAAGGTT
35 CTCGCCAAGGTGAAGAACAGCGCTACCCCTTAAGAGATCTCCTTGCCCTGGGTGGGAGGAG
CGAAAGTGGGGGTGTCTGGGGAGACCAGGAACCTGCCAAGCCCAGGCTGGGGCCAAGGA
CTCTGCTGAGAGGCCCTAGAGACAACACCCTTCCCAGGCCACTGGCTGCTGGACTGTTCT
CTCAGGAGCGGCCTGGGTACCCAGTATGTGCAGGGAGACGGAACCCCATGTGACAGCCCA
CTCCACCAGGGTTCCCAAAGAACCTGGCCCAGTCATAATCATTCATCCTGACAGTGGCAA
40 TAATCACGATAACCAGTACTAGCTGCCATGATCGTTAGCCTCATATTTTCTATCTAGAGC
TCTGTAGAGCAC

AA652388

5 GCGGCCGCAAGAAACGCATTCCGTACAGCAAGGGGCAGTTGCGGGACTGGAGCGTGAGTA
TGCGGCTAACAAGTTCATCACCAAGGACAAGAGGCGCAAGATCTCGGCAGCCACCAGCCT
CTCGGAGCGCCAGATTACCATCTGGTTTCAGAACCGCCGGGTCAAAGAGAAGAAGGTTCT
CGCCAAGGTGAAGAACAGCGCTACCCCTTAAGAGATCTCCTTGCTGGGTGGGAGGAGCG
AAAGTGGGGGTGTCCTGGGGAGACCAGGAACCTGCCAAGCCCAGGCTGGGGCCAAGGACT
CTGCTGAGAGGCCCCCTAGAGACAACACCCTTCCCAGGCCACTGGCTGCTGGACTGTTCT
10 CAGGAGCGGCCTGGGTACCCAGTATGTGCAGGGAGACGGAACCCCATGTGACAGCCCCT
CCACCAGGGTTCCCAAAGAACCTGGCC

BF446158

15 TTTTTTTTTTTTTTTTTTTTTTCGCCTTCCAAATTTATTCATAATTAGCTCAATTCATGAAAG
CGGTTTCTAAAGTGCTCTACAAAGCTCTAAATAAAAAATATGAGGCTAACGATCATGGCA
GCTAGTACTGGTTATCGGGATTATTGCCACTGTCAGGATGAATGATTATGACTGGGCCAG
GTTCTTTGGGAACCCTGGTGGAGTGGGCTGTCACATGGGGTTCCGTCTCCCTGCACATAC
TGGGTACCCAGGCCGTTCTGAGGAACAGTCCACCACCCAGTGGCCTGGGAAGGGTGTTG
TCTCTAGGGGCCTCTCAACAAAGTCCTTGGCCCCAGCCTGGGCTTGGCAGGTTCTTGCTC
20 TCCCCAGGACACCCCCACTTTCGCTCCTCCCACCCAGGCAAGGAGATCTCTTAAGGGG

AA657924

25 GACGCNAGGTATGCGGCTAACAAGTTCATCACCAAGGACAAGAGGCGCAAGATCTCGGCA
GCCACCAGCCTCTCGGAGCGCCAGATTACCATCTGGTTTCAGAACCGCCGGGTCAAAGAG
AAGAAGGTTCTCGCCAAGGTGAAGAACAGCGCTACCCCTTAAGAGATCTCCTTGCTGGG
TGGGAGGAGCGAAAGTGGGGGTGTCCTGGGGAGACCAGGAACCTGCCAAGCCCAGGCTGG
GGCCAAGGACTCTGCTGAGAGGCCCCCTAGAGACAACACCCTTCCCAGGCCACTGGCTGCT
GGACTGTTCTCAGGAGCGGCCTGGGTACCCATGTATGTGCAGGGAGACGGAACCCCATG
TGACAGCCCCTCCACCAGNGTTCCTAAAGAACCCTGGCCAGTCA

AA644637

30 GCAGGCGACTTGCGAGCTGGGAGCGGTTTAAAACGCTTTGGATTCCCCCGGCCTGGGTGG
GGAGAGCGAGCTGGGTGCCCCCTAGATTCCCCGCCCCCGCACCTCATGAGCCGACCCTCG
GTCCATGGACACGGCAATTATGCCACCTTGATGGAGCCAAGGATATCGAAGGCTTGCTG
GGAGCGGGAGGGGGGCGGAATCTGGTCGCCCACTCCCCTCTGACCAGCCACCCAGCGGCG
35 CCTACGCTGATGCCTGCTGTCAACTATGCCCCCTTGATCTGCCAGGCTCGGCGGACTCT
NAAAGCATATGCCACCCNATGCCCTGGGGTGCCCCAGGGGAACGTCCCCAGCTCCCGTGC
CTTATGGTT

BF222357

GCGGCCGCAAGAAACGCATTCCGTACAGCAAGGGGCAGTTGCGGGAGCTGGAGCGGGAGT
ATGCGGCTAACAAGTTCATCACCAAGGACAAGAGGCGCAAGATCTCGGCAGCCACCAGCC
TCTCGGAGCGCCAGATTACCATCTGGTTTCAGAACCGCCGGGTCAAAGAGAAGAAGGTTC
TCGCCAAGGTGAAGAACAGCGCTACCCCTTAAGAGATCTCCTTGCCTGGGTGGGAGGAGC
5 GAAAGTGGGGGTGTCCTGGGGAGACCAGGAACCTGCCAAGCCCAGGCTGGGGCCAAGGAC
TCTGCTGAGAGGCCCTTAGAGACAACACCCTTCCCAGGCCACTGGCTGCTGGACTGTTCC
TCAGGAGCGGCCTG

AA527613

10 GTCGACGAACAGCGCTACCCCTTAAGAGATCTCCTTGCCTGGGTGGGAGGAGCGAAAGTG
GGGGTGTCTTGGGGAGACCGGGAACCTGCCAAGCCCAGGCTGGGGCAAGGACTCTGCTGAG
AGGCCCCTAGAGACAACACCCTTCCCAGGCCACTGCTGCTGGACTGTTCTCAGGAGCGG
CCTGGGTACCCAGTATGTGCAGGGAGACGGAACCCCATGTGACAGCCCCTCCACCAGGG
TTCCCAAAGAACCTGGCCCAGTCATAATCATTCATCCTGACAGTGGCAATAATCACGATA
15 ACCAGTACTCAGCTGCCATGATCGTTAGCCTCATATT

AA533227

GCGTCGACCCCTTGAAGAGATCTCCTTGCCTGGGTGGGAGGAGCGAAAGTGGGGGTGTCC
TGGGGAGACCAGGAACCTGCCAAGCCCAGGCTGGGGCCAAGGACTCTGCTGAGAGGCCCC
20 TAGAGACAACACCCTTCCCAGGCCACTGGCTGCTGGACTGTTCTCAGGAGCGGCCTGGG
TACCCAGTATGTGCAGGGAGACGGAACCCCATGTGACAGCCCCTCCACCAGGGTTCCCA
AAGAACCTGGCCCAGTCATAATCATTCATCCTGACAGTGGCAATAATCACGATAACCAGT
ACTAGCTGCCATGATCGTTAGCCTCATATTTTCTATCTAGAGCTCTGTAGAGCACTTGTA
GAAACCGCTTTCATGAATTGAGCTAATTATGAATAGATTTGGAAGGGGAAAAAGTGGAA
25 AAAGTTTTGCCCAAAGTGGGTTCGTTTACGTCTG

AA456069

CTCCCTGGCAACACATCTGGCTGTTCCAGCACCAGCGAGACCCAAGACTGGTAACTGTCC
ACAGGCAACAGGGAGTCATGTGCGGTTTCTCCAGGAGCACCCAGAGTCTGCACCACAGAC
30 ACGTCCAGGTAACCTGGCCATAGCTGAGTAGGTTCCCGGATATCCCGGATAGAAGGCAAAC
TCAGTGGGGCGGCTGGGGTACTCTTCCCCGGCCGTGGAGAGTCTCCGCGGGGTACGGCCC
AGGGTGGCTGCCTGGGCATCAGGGTTTCAGCGAGCTCCGGGACACTCGGCAGGAGTAGTA
CCCGCCTCCAAAGTAACCATAAGGCACGGGAGCTGGGGACGTCCCTGGGGCACCCCAG

AA455572

TTTAAAACGCTTTGGATTCCCCCGGCCTGGGTGGGGAGAGCGAGCTGGGTGCCCCCTAGA
TTCCCCGCCCCCGCACCTCATGAGCCGACCCTCGGTCCATGGAGCCGGCGAATTATGCCA
CCTTGATGGAGCCAAGGATATCGAAGGCTTGCTGGGAGCGGGAGGGGGCGGAATCTGG
TCGCCCCTCCCTCTGACCAGCCACCCAGCGGCGTACGTGATGCCTGCTGTCAACTAT
40 GCCCTTGATCTGCCAGCTCGCGGAGCCAAAGCAATGCCACCCATGCCCTGGGGTGCCCC

AGGTGACGTCCCCAGCTCCCGTGCCTTATGGTTACTTTGGAGGCGGGTACTACTCCTGCC
GAGTGTCCCGGAGCTCGCTGAAACCCTGTGCCCAGGCAGCCACCCTGGCCGCGTACCCCG
CGATGACTCCCACGGCCGGGAAGAGTACCCAGCCGCCCCACTGAGTTTGCCT

5 BX117624

CAGGCGACTTGCGAGTCTGGGAGCGATTTAAAACGCTTTGGATTCCCCCGGCCTGGGTGG
GGAGAGCGAGCTGGGTGCCCCCTAGATTCCCCGCCCCCGCACCTCATGAGCCGACCCTCG
GCTCCATGGAGCCCGCAATTATGCCACCTTGGATGGAGCCAAGGATATCGAAGGCTTGC
TGGGAGCGGGAGGGGGGCGGAATCTGGTCGCCCCACTCCCCTCTGACCAGCCACCCAGCGG
10 CGCCTACGCTGATGCCTGCTGTCAACTATGCCCCCTTGGATCTGCCAGGCTCGGCGGAGC
CGCCAAAGCAATGCCACCCATGCCCTGGGGTGCCCCAGGGGACGTCCCCAGCTCCCGTGC
CTTATGGTTACTTTGGAGGCGGGTACTACTCCTGCCGAGTGTCCCGGAGCTCGCTGAAAC
CCTGTGCCCAGGCAGCCACCCTGGCCGCGTACCCCGCGGAGACTCCCACGGCCGGGGAAG
AGTACCCAGCCGCCCCACTGAGTTTGCCTTCTATCCGGGATATCCGGGAACCTACCAGC
15 CTATGGCCAGTTACCTTGGACGTGTCTGTGGTGCAGACTCTGGGTGCTCCTGGAGAACCG
CGACATGACTCCCTGNTGCCTGTGGACAGTTACCAGTCTTGGGCTCTCGCTGGTGGCTGG
AACAGCCAGATGTGTTGNCAGGGAGAACAGAACCCACCAGGTCCCTTTTGAAGGCAGAT
TTGCAGACTNCAGCGGGCA

20 BQ673782

AGGCAGCCACCCTGGCCGCGTACCCCGCGGAGACTCCACGGCCGGGGAAGAGTACCCCA
GCCGCCCCACTGAGTTTGCCTTCTATCCGGGATATCCGGGAACCTACCAGCCTATGGCCA
GTTACCTGGACGTGTCTGTGGTGCAGACTCTGGGTGCTCCTGGAGAACCGCGACATGACT
CCCTGTTGCCTGTGGACAGTTACCAGTCTTGGGCTCTCGCTGGTGGCTGGAACAGCCAGA
25 TGTGTTGCCAGGGAGAACAGAACCCACCAGGTCCCTTTTGAAGGCAGCATTTGCAGACT
CCAGCGGGCAGCACCTCCTGACGCCTGCGCCTTTCGTGCGCGCCGCAAGAAACGCATTC
CGTACAGCAAGGGGCAGTTGCGGGAGCTGGAGCGGGAGTATGCGGCTAACAAGTTCATCA
CCAAGGACAAGAGGCGCAAGATCTCGGCAGCCACCAGCCTCTCGGAGCGCCAGATTACCA
TCTGGTTTTCAGAACCGCCGGGTCAAAGAGAAGAAGTTCTCGCCAAGGTGAAGAACAGCG
30 CTACCCCTTAAGAGATCTCCTTGCCTGGGTGGGAGGAGCGAAAGTGGGGGTGTCCTGGGG
AGACCAGGAACCTGCCAAGCCCCAGGCTGGGGCCAAGGACTCTGCTGAGAGGCCCTTAGA
GACAACACCCTTCCCAGGCCACTGGCTGCTGGACTGTTCCCTCAGGAGCGGCCTGAGTACC
CCGTATGTGCAGGGGAGACGGAACCCCTGTGACCAGCCCCCTCCACCCGTGGTCTCCC
AGATAACCTGGCCCCCACTCATAAATCATTTCTTCCCGGGCCGGGGGCCAATCATTTCCC
35 GAACTACCCCGGTACCTTATACAATTAGATTGGACATGAATCCTCTCGGGGGCATTCCTT
ATGGCGCTGAGGCCCTCACACCT

AI814453

GGGTGCTGTCCTCTGGAGTCTGCAAATGCTGCCTTCCAAAAGGGACCTGGTGGGTCTGT
40 TCTCCCTGGCAACACATCTGGCTGTTCCAGCCACCAGCGAGAGCCCAAGACTGGTAACCTG
TCCACAGGCAACAGGGAGTCATGTCGCGGTTCTCCAGGAGCACCCAGAGTCTGCACCACA

GACACGTCCAGGTAAC TGGCCATAGGCTGGTAGGTTCCCGGATATCCCGGATAGAAGGCA
AACTCAATGGGGCGGCTGGGGTACTCTTCCCCGGCCGTGGGAGTCTCCGCGGGGTACGCG
GCCAGGGTGGCTGCCTGGGCACAGGGTTTCAGCGAGCTCCGGGACACTCGGCAGGAGTAG
TACCCGCCTCCAAAGTAACCATAAGGCACGGGAGCTGGGGACGTCCCCTGGGGCACCCCA
5 NGGCATGGGTGGCATTGCTTTGGCGGCTCCGCCGAGCCTGGCAGATCCAAGGGGGCATAG
TTGACAGCAGGCATCAGCGTAGGCGCCGCTGGGTGGCTGGTCAAAGGGAGTGGCGACCA
NATTCCGCCCCCTCCCGCTTCCAG

AI417272

10 GGGTGCTGCCCCGCTGGAGTCTGCAAATGCTGCCTTCCAAAAGGGACCTGGTGGGTCTGT
TCTCCCTGGCAACACATCTGGCTGTTCCAGCCACCAGCGAGAGCCCAGGACTGGTAACTG
TCCACAGGCAACAGGGAGTCATGTCGCGGTTCTCCAGGAGCACCCAGAGTCTGCACCACA
GACACGTCCAGGTAAC TGGCCATAGGCTGGTAGGTTCCCGGATATCCCGGATAGAAGGCA
AACTCAGTGGGGCGGCTGGGGTACTCTTCCCCGCCGTGGGAGTCTCCGCGGGGTACGCGG
15 CCAGGGTGGCTGCCTGGGCACAGGGTTTCAGCGAGCTCCGGGACACTCGGCAGGAGTAGT
ACCCGCCTCCAAAGTAACCATAAGGCACGGGAGCTGGGGACGTCCCCTGGGGCACCCAG
GGCATGGGTGGCATTGCTTTGGCGGCTCCGCCGAGCCTGGCAGATCCAAGNGGCATAGT
TGACAGCAGGCATCAGCGTANGCGCCGCTGGGTGGCTGTCAAGAGG

20 AA535663

TCGACGTTACCTGGACGTGTCTGTGGTGCAGACTCTGGGTGCTCCTGGAGAACCGCGACA
TGACTCCCTGTTGCCTGTGGACAGTTACCAGTCTTGGGCTCTCGCTGGTGGCTGGAACAG
CAGATGTGTTGCCAGGGAGAACAGAACCCACCAGGTCCCTTTTGGAAGGCAGCATTTGCA
GACTCCAGCGGGCAGCACCCCTCCTGACGCCTGCGCCTTTCGTGCGGGCCGCAAGAAACGC
25 ATTCCGTACAGCAAGGGGCAGTTGCGGGACTGGAGCGGGAGTATGCGGCTAACAAGTTCA
TCACCAAGGACAAGAGGCGCAAGATCTCGGCAGCCACCAGCCTCTCGGAGCGCCAGATTA
CCATCTGGTTTTAGAACCGCCGGGTCAAAGAGAAGAAGGTTCTCGCCAAGGTGAAGAACA
GCGCTACCCCTTAAGAGATCTCCTTGCCTGGGTGGGAGGAGCGAAAGTGTG

30 AI400493

GTCAGGAGGGTGCTGCCCCGCTGGAGTCTGCAAATGCTGCCTTCCAAAAGGGACCTGGTGG
GTTCTGTTCTCCCTGGCAACACATCTGGCTGTTCCAGCCACCAGCGAGAGCCCAGGACTG
GTAAGTGTCCACAGGCAACAGGGAGTCATGTCGCGGTTCTCCAGGAGCACCCAGAGTCTG
CACCACAGACACGTCCAGGTAAC TGGCCATAGGCTGGTAGGTTCCCGGATATCCCGGATA
35 GAAGGCAAAC TCAAGTGGGGCGGCTGGGGTACTCTTCCCCGGCCGTGGGAGTCTCCGCGGG
GTACGCGGCCAGGGTGGCTGCCTGGGCACAGGGTTTCAGCGAGCTCCGGGACACTCGGCA
TGAGTAGACCCGCCTTCCAAGTAACCATAAGGCACGGGAGCTGGTAACGTCCCCTGGGGC
ACCCANGGCCATGGGTGCATTGCTTTGGCGGCTCCGCCGAGCCCTGCAGATCCAAGGTG
GGCATATTGACAGCAGGCATTCAGTATGCGCCCCCTGGGTGGCTGTCATATTGGGGATT
40 GCGAC

AW779219

GCAGGCGTCAGGAGGGTGCTGCCCCGCTGGAGTCTGCAAATGCTGCCTTCCAAAAGGGACC
TGGTGGGTTCTGTTCTCCCTGGCAACACATCTGGCTGTTCCAGCCACCAGCGAGAGCCCA
AGACTGGTAACTGTCCACAGGCAACAGGGAGTCATGTGCGGGTTCTCCAGGAGCACCCAG
5 AGTCTGCACCACAGACACGTCCAGGTAAGTGGCCATAGGCTGGTAGGTTCCCGGATATCC
CGGATAGAAGGCAAACCTCAGTGGGGCGACTGGGGTACTCTTCCCGGCCGTGGGGAGTCTC
CGCGGGGTACGCGGCCAGGGGTGGCTGCCTGGGCACCAGGGGTTTCAGCGAGCTCCGGGA
CACTCNGCAGGAAANTAGTACCCGCCTCCCAAAGTAACCATAAGCACCGGACTGNGGGNN
GGACGTCCCCTGGGGCAC

AA594847

GCGACCGGACGAAAGGAGGCGTCAGGAGGGTGCTGCCCCGCTGGAGTCTGCAAATGCTGCC
TTCCAAAAGGGACCTGGTGGGTTCTGTTCTCCCTGGCAACACATCTGGCTGTTCCAGCAC
CAGCGAGACCCAAGACTGGTAACTGTCCACAGGCAACAGGGAGTCATGTGCGGGTTCTCC
15 AGGAGCACCCAGAGTCTGCACCACAGACACGTCCAGGTAAGTGGCCATAGCTAGGTAGGT
TCCCGGATATCCCGGATAGAAGGCAAACCTCAGTGGGGCGACTGGGGTACTCTTCCCGGC
CGTGGGAGTCTCCGCGGGGTACGCCCATGGGTGGCTGCCTGGGCACAGGGTTTCAGCGAG
CTCCGGGACA

AI150430

GCAGGCGTCAGGAGGGTGCTGCCCCGCTGGAGTCTGCAAATGCTGCCTTCCAAAAGGGACC
TGGTGGGTTCTGTTCTCCCTGGCAACACATCTGGCTGTTCCAGCCACCAGCGAGAGCCCA
AGACTGGTAACTGTCCACAGGCAACAGGGAGTCATGTGCGGGTTCTCCAGGAGCACCCAG
AGTCTGCACCACAGACACGTCCAGGTAAGTGGCCATAGGCTGGTAGGTTCCCGGATATCC
25 CGGATAGAAGGCAAACCTCAGTGGGGCGACTGGGGTACTCTTCCCGGCCGTGGGAGTCTC
CGCGGGGTACGCGGCCAGGGTGGCTGCCTGGGCACAGGGTTTCAGCGAGCTCCGGGACAC
TCGGCAGGAGTAGTACCCGCCTCCAAAGTAACCATAAGGCACGGGAGCTGGATGCGTCCC
CTAGGGCACCCCATGGCATGGGTGGCATTGCTTTGGCGGCTCCGCCGAGCCTGGCAGATC
CAAGGAGGCACTGTT

AA494387

GGGTGCTGCCCCGCTGGAGTCTGCAAATGCTGCCTTCCAAAAGGGACCTGGTGGGTTCTGT
TCTCCCTGGCAACACATCTGGCTGTTCCAGCCACCAGCGAGACCCAAGACTGGTAACTGT
CCACAGGCAACAGGGAGTCATGTGCGGGTTCTCCAGGAGCACCCAGAGTCTGCACCACAG
35 ACACGTCCAGGTAAGTGGCCATAGGCTGGTAGGTTCCCGGATATCCCGGATAGAAGGCAA
ACTCAGTGGGGCGGCTGGGGTACTCTTCCCGGCCGTGGGAGTCTCCGCGGGGTACGCGT
CCAGGGTGGCTGCCTGGGCACAGGGTTTCAGCGAGCTCCGGGACACTCGGCAGGAGTAGT
ACCCGCCTCCAAAGTAACCATAAGGCACGGGAGCTGGGGACGTCCCTG

AA662643

GGGTGCTGCCCCGCTGGAGTCTGCAAATGCTGCCTTCCAAAAGGGACCTGGTGGGTTCTGT
TCTCCCTGGCAACACATCTGGCTGTTCCAGCCACCAGCGAGACCCAAGACTGGTAACTGT
CCACAGGCAACAGGGAGTCATGTGCGGGTTCTCCAGGAGCACCCAGAGTCTGCACCACAG
ACACGTCCAGGTAACCTGGCCATAGGTGGTAGGTTCCCGGATATCCCGGATAGAAGGCAAA
5 CTCAGTGGGGCGGCTGGGGTACTCTTCCCCGGCCGTGGGAGTCTCCGCGGGGTACGCGGC
CAGGGTGGCTGCCTGGGCACAGGGTTTCAGCGAGCTCCGGGACA

AI935940

GGGTGCTGCCCCGCTGGAGTCTGCAAATGCTGCCTTCCAAAAGGGACCTGGTGGGTTCTGT
10 TCTCCCTGGCAACACATCTGGCTGTTCCAGCCACCAGCGAGAGCCCAAGACTGGTAACTG
TCCACAGGCAACAGGGAGTCATGTGCGGGTTCTCCAGGAGCACCCAGAGTCTGCACCACA
GACACGTCCAGGTAACCTGGCCATAGGCTGGTAGGTTCCCGGATATCCCGGATAGAAGGCA
AACTCAGTGGGGCGGCTGGGGTACTCTTCCCCGGCCGTGGGAGTCTCCGCGGGGTACGCG
15 GCCAGGGTGGCTGCCTGGGCACAGGGTTTCAGCG

AA532530

GGGTGCTGCCCCGCTGGAGTCTGCAAATGCTGCCTTCCAAAAGGGACCTGGTGGGTTCTGT
TCTCCCTGGCAACACATCTGGCTGTTCCAGCCACCAGCGAGACCCAAGACTGGTAACTGT
CCACAGGCAACAGGGAGTCATGTGCGGGTTCTCCAGGAGCACCCAGAGTCTGCACCACAG
20 ACACGTCCAGGTAACCTGGCCATAGGTNGGTAGGTTCCCGGATATCCCGGATAGAAGGCAA
ACTCAGTGGGGCGGCTGGGGTACTCTTCCCCGGCCGTGGGAGTCTCCG

AA857572

CTCCCTGGCAACACATCTGGCTGTTCCAGCACCAGCGAGAGCCAAGACTGGTAACTGTCC
25 ACAGGCAACAGGGAGTCATGTGCGGGTTCTCCAGGAGCACCCAGAGTCTGCACCACAGAC
ACGTCCAGGTAACCTGGCCATAGGTGCGGTAGGTTCCCGGATATCCCGGATAGAAGGCAAA
TCAGTGGGGCGACTGGGGTACTCTTCCCCGGCCGTGGGAGTCTCCGCGGGGTACGGCNAC
AGGGTGGCTGCCTGGGCACAGGGTTTCAGCGAGCTCCGGGACACTCGGCAGGAGTAGTAN
30 CCGCCTCAAAGTAACCATAANGCACGGGAGCTGGGGACGTCCC

AI261980

ACGAAAGGCGCAGGCGTCAGGAGGGTGCTGCCCCGCTGGAGTCTGCAAATGCTGCCTTCCA
AAAGGGACCTGGTGGGTTCTGTTCTCCCTGGCAACACATCTGGCTGTTCCAGCCACCAGC
GAGAGCCCAAGACTGGTAACTGTCCACAGGCAACAGGGAGTCATGTGCGGGTTCTCCAGG
35 AGCACCCAGAGTCTGCACCACAGACACGTCCAGGTAACCTGGCCATAGGCTGGTAGGTTCC
CGGATATCCCGGATAGAAGGCAAACCTCAGTGGGGCGACTGGGGTACTCTTCCCCGGCCCCG
GGGAGTCTCCGCGGGGTACGCGGCCAGGGTGGCTGCCTGGGCACAGGGTTTCAGCGAGCT
CCGGGACACTCGGCGGAGNTAGTACCCGCCTCCAAAGTAACCATAAGGCACGGGAGCTGG
40 GGAACCGTCCCCTGGGGCACC

BE888751.1

GAGCGAGCTGGGTGCCCCCTAGATTCCCCGCCCCCGCACCTCATGAGCCGACCCTCGGCT
CCATGGAGCCCGGCAATTATGCCACCTTGGATGGAGCCAAGGATATCGAAGGCTTGCTGG
GAGCGGGAGGGGGGCGGAATCTGGTCGCCCCTCCCTCTGACCAGCCACCCAGCGGCGC
5 CTACGCTGATGCCTGCTGTCAACTATGCCCCCTTGGATCTGCCAGGCTCGGCGGAGCCGC
CAAAGCAATGCCACCCATGCCCTGGGGTGCCCCAGGGACGTCCCCAGCTCCCGTGCCTTA
TGGTTACTTTGGAGGCGGGTACTACTCCTGCCGAGTGTCCCGGAGCTCGCTGAAACCCTG
TGCCCAGGCAGCCACCCTGGCCGCGTACCCCGCGGAGACTCCACGGCCGGGGAAGAGTA
CCCCAGCCGCCCCACTGAGTTTGCTTCTATCCGGGATATCCGGGAACCTACCAGCCTAT
10 GGCCAGTTACCTGGACGTGTCTGTGGTGCAGACTCTGGGTGCTCCTGGAGAACC GCGACA
TGA CTCCCTGTTGCCTGTGGACAGTTACCAGTCTTGGGCTCTCGCTGGTGGCTGGAACAG
CCAGATGTGTTGCCAGGGAGAACAGAACCCACCAGGTCCCTTTTTTGAAGGCAGCATTTG
CAGACTCCAGCGGCAGGACCTCCTGAACGCCTGCGCCTTTCGTGCGGCGTCTAAAGTAA
TCCTCGAGG

15

AI378797

GCGGCCGCGGCCACCACCAACTGCTCGCCACCGACCCCACTACTCGCCACCGACCCGCT
GCTCGGAGCTTCGGTTCTGCGGGTTGTCCAGACTTCAGGCCTGTGCGCTCAATCGTGGAG
AATGCGCCGGCAGGCCCCCACCAGCCTAAGGTGCAGGAAGGACCAGCACGAACCCG
20 CTGGCTTTGCTGCGCGGCCAGGAGATGAGTCCACCGGGCACTGAGCCCAGGTACAGGAC
ATCAGAGAATGAACACAGAGGCAGAGGCCCTCATGTCCCTCTCAGAGTCCCGGCTCTGCA
NAGAGCCCGTCTGTCTCCAGCTTCCAGAATTCCGCACTGTGAATCTGTCTACGTGGACTG
GGAAAACAGGGTTGGCACCCTCTGCCACTCCGTTTGTGCCTGGGAAGGGCTAAGTATGC
AAGGCTACAAACATCTACTTCACTGGGATCCCAAATGCTCAACAAACCATGACCTGCTNT
25 GGTCAGAACCACCAGAAATATT

AA234220

GCAGGCGACTTGCGAGCTGGGAGCACTTTAAACGCTTTGGATTCCCCCGGCCTGGGTGG
GGAGAGCGAGCTGGGTGCCCCCTAGATTCCCCGCCCCCGCACCTCATGAGCCGACCCTCG
30 GCTCCATGGAGCCTGGCATATTATGCCACCTTGGTATGGAGCCAAGGATATCGAAGGCTT
GCTGGGAGCGGGAGGGGGGCGGAATCTGGTCGCCCCTCCCTCTGACCAGCCACCCAGC
GGCGCCTACGCTGATGCCTGCTGTCAACTATGCCCCCTTGA

AA236353

GCCCCGCTGGAGTCTGCAAATGCTGCCTTCCAAAAGGGACCTGGTGGGTCTGTTCTCCCT
GGCAACACATCTGGCTGTTCCAGCCACCAGCGAGACGCCAAGACTGGTAACTGTCCACAG
GCAACAGGGAGTCATGTCGCGGTTCTCCAGGAGCACCCAGAGTCTGCACCACAGACACGT
CCAGGTAAC TGCCATAGGTNGGTAGGTTCCCGGATATCCCGGATAGAAGGCAAACCTCAG
TGGGGCGGCTGGGGTACTCTTCCCCGGCCGTGGGAGTCTCCGCGGGGTACGCGCACAGGG
40 TGGCTGCCTGGGCACAGGGTTTCAGCGAGCTCCGGGACACTCGGCAGGAGTAGTACCCGC
CTCCAAAGTAACCATAAGGCA

AA588193

AACTGCTCGCCACCGACCCCACTACTCGCCACCGACCCGCTGCTCGGAGCTTCGGTTCTG
CGGGTTGTCCAGACTTCAGGCCTGTGCGCTCAATCGTGGAGAATGCGCCGGCAGCCCCCA
5 CCCCCAGCCTAAGGTGCAGGAAGGACCAGCACGAACCCGCTGGCTTTGCTGCGCGGCCAG
GAGATGAGTCCCACCGGGCACTGAGCCCAGGTACAGGACATCAGAGAATGAACACAGAGG
CAGAGGCCCTCATGTCCCTCTCAGAGTCCCGGCTCTGCAAAGAGCCCGTCTGTCTCCAGC
TTCCAGAATTCGCACTGTGAATCTGTCTACGTGGACTGGGAAAACAGGGTTGGCACCAC
TCTGCCACTCCGTTTGTGCCTGGGAAGGGCTAAGTATGCAAGGCT

10

AI821103

GATCCCTTTGCAGGGAAGCTTTCTCTCAGACCCCCTTCCATTACACCTCTCACCCCTGGTA
ACAGCAGGAAGACTGAGGAGAGGGGAACGGGCAGATTTCGTTGTGTGGCTGTGATGTCCGT
TTAGCATTTTTCTCAGCTGACAGCTGGGTAGGTGGACAATTGTAGAGGCTGTCTCTTCCT
15 CCCTCCTTGTCCACCCCATAGGGTGTACCCACTGGTCTTGGAAGCACCCATCCTTAATAC
GATGATTTTTCTGTCTGTGAAAATGAAGCCAGCAGGCTGCCCCTAGTCAGTCCTTCCTT
CCAGAGAAAAAGAGATTTGAGAAAGTGA

AI821851

TTTTTTTTTTTTTTTTTTTTTTTTCTTTTTCACTTTCTCAAATCTCTTTTTCTCTGGAAGGAAG
GACTGACTAGGGGCAGCCTGCTGGCTTCATTTTACACGACAAAAAATCATCGTATTAA
GGATGGGTGCTTCCAAAACCAAGTGGGTACACCCTATGGGGGGGACAAGGAGGGAGGAAGA
GACAGCCTCTACAATTGTCCACCTACCCAGCTGTCAGCTGAGAAAAATGCTAAACGGACA
TCACAGCCACACAACGAATCTGCCCGTTCCCCTCTCCTCAGTCTTCCTGCTGTTACCAGG
25 GTGAGAGGTGTAATGGAAGG

AA635855

TTTTTTTTTTTTTTTTTTTTTTTTCTTTTTCACTTTCCCAAATCTCTTTTTCTCTGGAAGGAAG
GACTGACTAGGGGCAGCCTGCTGGCTTCATTTTACACGACAGAAAAATCATCGTATTAA
30 GGATGGGTGCTTCCAAGACCAAGTGGGTACACCCTATGGGGTGGACACAGGAGGGAGGAAG
AGACAGCCTCTACAATTGTCCACCTACCCAGCTGTCAGCTGAGAAAAATGCTAAACGGAC
ATCACAGCCACACAACGAATCTGCCCGTTCCCCTCTCCTCAGTCTTCCTGCTGTTACCAG
GGTGAGAGGTGTAATGGAAGG

AI420753

GCGGCCGCGGCCACCACTGCTCGCCACCGACCCCACTACTCGCCACCGACCCGCT
GCTCGGAGCTTCGGTTCTGCGGGTTGTCCAGACTTCAGGCCTGTGCGCTCAATCTTGAG
AATGCGCCGGCAGGCCCCCACCACCCAGCCTAAGGTGCAGGAAGGACCAGCACGAACCCG
CTGGCTTTGCTGCGCGGCCAGGAGATGAGTCCACCGGGCACTGAGCCCAGGTACAGGAC

ATCAGAGAATGAACACAGAGGCAGAGGCCCTCATGTCCCTCTCAGAGTCCCGGCTCTGCA
AAGAGCCCGTCTGTCTCCAGCTTCCAGAATTCCGCACTGTGAATCTGTCTACGT

BG180547

5 CACGCGTCGATCCCAGTGAAGTAGATGTTTGTAGCCTTGCATACTTAGTCCTTCCCAGGC
ACAAACGGAGTGGCAGAGTGGTGCCAACCCTGTTTTCCCAGTCCACGTAGACAGATTAC
AGTGCGGAATTCTGGAAGCTGGAGACAGACGGGCTCTTTGCAGAGCCGGGACTCTGAGAG
GGACATGAGGGCCTCTGCCTCTGTGTTTATTCTCTGATGTCCTGTACCTGGGCTCAGTGC
10 CCGGTGGGACTCATCTCCTGGCCGCGCAGCAAAGCCAGCGGGTTCGTGCTGGTCTTCTTCT
GCACCTTAGGCTGGGGGTGGGGGGCCTGCCGGCGCATTCTCCACGATTGAGCGCACAGGC
CTGAAGTCTGGACAACCCGCAGAACCGAAGCTCCGAGCAGCGGGTCGGTGGCGAGTAGTG
GGGTCGGTGGCGAGCAGTTGGTGGTGGG

AA468306

15 TCGACCTCGCCAAGGTGAAGAACAACGCTACCCCTTAAGAGATCTCCTTGCCTGGGTGGG
AGGAGCGAAAGTGGGGGTGTCCTGGGGAGACCAGGAACCTGCCAAGCCCAGGCTGGGGCC
AAGGACTCTGCTGAGAGGGCCCTAGAGACAACACCCTTCCCAGGCCACTGGCTGCTGGAC
TGTTCTCTCAGGAGCGGCCTGGGTACCCAGTATGTGCAGGGAGA

20 **AA468232**

TTTTTTACTGGTTATCGTGGTTATTGCCACTGTCAGGATGAATGATTATGACTGGGCCAG
GTTCTTTGGGAACCCTGGTGGAGTGGGCTGTCACATGGGGTTCCGTCTCCCTGCACATAC
TGGGTACCCAGGCCGCTCCTGAGGAACAGTCCAGCAG

25 **CB050115**

GGCCACCACTGCTCGCCACCGACCCCACTACTCGCCACCGACCCGCTGCTCGGAG
CTTCGGTTCTGCGGGTTGTCCAGACTTCAGGCCTGTGCGCTCAATCGTGGAGAATGCGCC
GGCAGGCCCCCAGCCCCAGCCTAAGGTGCAGGAAGGACCAGCACGAACCCGCTGGCTTT
GCTGCGCGGCCAGGAGATGAGTCCACCGGGCACTGAGCCCAGGTACAGGACATCAGAGA
30 ATGAACACAGAGGCAGAGGCCCTCATGTCCCTCTCAGAGTCCCGGCTCTGCAAAGAGCCC
GTCTGTCTCCAGCTTCCAGAATTCCGCACTGTGAACCTCGTGCC

CB050116

GGCACGAGGTTACAGTGCGGAATTCTGGAAGCTGGAGACAGACGGGCTCTTTGCAGAGC
35 CGGGACTCTGAGAGGGACATGAGGGCCTCTGCCTCTGTGTTTATTCTCTGATGTCCTGTA
CCTGGGCTCAGTGCCCGGTGGGACTCATCTCCTGGCCGCGCAGCAAAGCCAGCGGGTTCC
TGCTGGTCTTCTCTGCACCTTAGGCTGGGGGTGGGGGGCCTGCCGGCGCATTCTCCACGA
TTGAGCGCACAGGCCTGAAGTCTGGACAACCCGCAGAACCGAAGCTCCGAGCAGCGGGTC
GGTGGCGAGTAGTGGGGTTCGGTGGCGAGCAGTTGGTGGTGGGCC

AA661819

GCTGCTCGGAGCTTCGGTTCTGCGGGTTGTCCAGACTTCAGGCCTGTGCGCTCAATCGTG
GAGAATGCGCCGGCAGCCCCCAGCCTAAGGTGCAGGAAGGACCAGCACGAACCC
5 GCTGGCTTTGCTGCGCGGCCAGGAGATGAGTCCCACCGGCACTGAGCCAGGTACAGGACA
TCAGAGAATGAACACAGAGGCAGAGGCCTCATGTCCCTCTCAGAGTCCCGGCTCTGCAAA
GAGCCGTACTGTCTCCAGCTTCCAGAATTCCGCACTGTGAATCTGTCTACGTGGACTGGG
AAAAC

10 CF146837

CACGAGGATTTTCTATCTAGAGCTCTGTAGAGCACTTTAGAAACCGCTTTCATGAATTGA
GCTAATTATGAATAAATTTGGAAGGCGATCCCTTTGCAGGGAAGCTTTCTCTCAGACCCC
CTTCCATTACACCTCTCACCTTGGTAACAGCAGGAAGACTGAGGAGAGGGGAACGGGCAG
ATTCGTTGTGTGGCTGTGATGTCCGTTTAGCATTTTTCTCAGCTGACAGCTGGGTAGGTG
15 GACAATTGTAGAGGCTGTCTCTTCCCTCCCTCCTTGTCCACCCCATAGGGTGTACCCACTG
GTCTTGGAAACACCCATCCTTAATACGATGATTTTTCTGTCGTGTGAAAATGAAGCCAGC
AGGCTGCCCCCTAGTCAGTCCTTCCTTCCAGAGAAAAAGAGATTTGAGAAAGTGCCTGGGT
AATTCACCATTAATTTCTCCCCCAAACCTCTCTGAGTCTTCCCTTAATATTTCTGGTGGT
TCTGACCAAAGCAGGTCATGGTTTGTGTGAGCATTTGGGATCCCAGTGAAGTAGATGTTTG
20 TAGCCTTGCATACTTAGCCCTTCCCAGGCACAAACGGAGTGGCAGAGTGGTGCCAACCCCT
GTTTTCCAGTCCACGTAGACAGATTACAGTGCGGAATTCTGGAAGCTGGAGACAGACG
GGCTCTTTGCAGAGCCGGGACTCTGAG

CF146763

25 CACGAGGATTTTCTATNCTAGAGCTCTGGTAGAGCACTTTANAAACCGCTTTCATGAATT
GAGCTAATTATGAATAAATTTGGAAGGCGATCCCTTTGCAGGGAAGCTTTCTCTCAGACC
CCCTTCCATTACACCTCTCACCTTGGTAACAGCAGGAAGACTGAGGAGAGGGGAACGGGC
AGATTCGTTGTGTGGCTGTGATGTCCGTTTAGCATTTTTCTCAGCTGACAGCTGGGTAGG
TGGACAATTGTAGAGGCTGTCTCTTCCCTCCCTCCTTGTCCACCCCATAGGGTGTACCCAC
30 TGGTCTTGGAAACACCCATCCTTAATACGATGATTTTTCTGTCGTGTGAAAATGAAGCCA
GCAGGCTGCCCCCTAGTCAGTCCTTCCTTCCAGAGAAAAAGAGATTGAGAAAGTGCCTGGG
TAATTCACCATTAATTTCTCCCCCAAACCTCTCTGAGTCTTCCCTTAATATTTCTGGTGG
TTCTGACCAAAGCAGGTCATGGTTTGTGTGAGCATTTGGGATCCCAGTGAAGTAGATGTTT
GTAGCCTTGCATACTTAGCCCTTCCCAGGCACAAACGGAGTGGCAGAGTGGTGCCAACCC
35 TGTTTTCCAGTCCACGTAGACAGATTACAGTGCGGAATTCTGGAAGCTGGAGACAGAC
GGGCTCTTTGCAGAGCCGGGACTCTGA

CF144902

40 CACGAGGGAAGCCAGCAGGCTGCCCCCTAGTCAGTCCTTCCTTCCAGAGAAAAAGAGATTT
GAGAAAGTGCCTGGGTAATTCACCATTAATTTCTCCCCCAAACCTCTCTGAGTCTTCCCT

TAATATTTCTGGTGGTTCTGACCAAAGCAGGTCATGGTTTGTTGAGCATTTGGGATCCCA
GTGAAGTAGATGTTTGTAGCCTTGCATACTTAGCCCTTCCCAGGCACAAACGGAGTGGCA
GAGTGGTGCCAACCCTGTTTTCCCAGTCCACGTAGACAGATTCACAGTGCGGAATTCTGG
AAGCTGGAGACAGACGGGCTCTTTGCAGAGCCGGGACTCTGAGAGGGACATGAGGGCCTC
5 TGCCTCTGTGTTTCACTCTCTGATGTCCTGTACCTGGGCTCAGTGCCCGGTGGGACTCATC
TCCTGGGCGCGCAGCAAAGCCAGCGGGTTCGTGCTGGTCCTTCCTGCACCTTA

CF141511.1

CACGAGGCCCTGGTAACAGCAGGAAGACTGAGGAGAGGGGAACGGGCAGATTCGTTGTGTG
10 GCTGTGATGTCCGTTTAGCATTTTTCTCAGCTGACAGCTGGGTAGGTGGACAATTGTAGA
GGCTGTCTCTTCCCTCCCTCCTTGTCCACCCCATAGGGTGTACCCACTGGTCTTGAAACA
CCCATCCTTAATACGATGATTTTTCTGTCTGTGAAAATGAAGCCAGCAGGCTGCCCTA
GTCAGTCCTTCCCTCCAGAGAAAAAGAGATTTGAGAAAGTGCCTGGGTAAATTCACCATTA
ATTTCCCTCCCCCAAACCTCTCTGAGTCTTCCCTTAATATTTCTGGTGGTTCTGACCAAAGC
15 AGGTCATGGTTTGTTGAGCATTTGGGATCCCAGTGAAGTAGATGTTTGTAGCCTTGCATA
CTTAGCCCTTCCCAGGCACAAACGGAGTGGCAGAGTGGTGCCAACCCTGTTTTCCCAGTC
CACGTAGACAGATTCACAGTGCGGAATTCTGGAA

CF139563.1

CACGAGGTCTTCCCTTAATATTTCTGGTGGTTCTGACCAAAGCAGGTCATGGTTTGTTGA
GCATTTGGGATCCCAGTGAAGTAGATGTTTGTAGCCTTGCATACTTAGCCCTTCCCAGGC
ACAAACGGAGTGGCAGAGTGGTGCCAACCCTGTTTTCCCAGTCCACGTAGACAGATTCAC
AGTGCGGAATTCTGGAAGCTGGAGACAGACGGGCTCTTTGCAGAGCCGGGACTCTGAGAG
GGACATGAGGGCCTCTGCCTCTGTGTTTCACTCTCTGATGTCCTGTACCTGGGCTCAGTGC
25 CCGGTGGGACTCATCTCCTGGCCGCGCAGCAAAGCCAGCGGGTTCGTGCTGGTCCTTCCT
GCACCTTAGGCTGGGGGTGGGGGGCCTGCCGGCGCATTCTCCACGATTGAGCGCACAGGC
CTGAAGTCTGGACAACCCGCGAGAACCGAAGCTCCGAGCAGCGGGTCGGTGGCGAGTA

CF139372

CACGAGGATTTCTGGTGGTTCTGACCAAAGCAGGTCATGGTTTGTTGAGCATTTGGGATC
CCAGTGAAGTAGATGTTTGTAGCCTTGCATACTTAGCCCTTCCCAGGCACAAACGGAGTG
GCAGAGTGGTGCCAACCCTGTTTTCCCAGTCCACGTAGACAGATTCACAGTGCGGAATTCT
TGGAAGCTGGAGACAGACGGGCTCTTTGCAGAGCCGGGACTCTGAGAGGGACATGAGGGC
CTCTGCCTCTGTGTTTCACTCTCTGATGTCCTGTACCTGGGCTCAGTGCCCGGTGGGACTC
35 ATCTCCTGGCCGCGCAGCAAAGCCAGCGGGTTCGTGCTGGTCCTTCCTGCACCTT

CF139319

CACGAGGAAGGCGATCCCTTTGCAGGGAAGCTTTCTCTCAGACCCCTTCCATTACACCT
CTCACCTGGTAACAGCAGGAAGACTGAGGAGAGGGGAACGGGCAGATTCGTTGTGTGGC
40 TGTGATGTCCGTTTAGCATTTTTCTCAGCTGACAGCTGGGTAGGTGGACAATTGTAGAGG

CTGTCTCTTCCTCCCTCCTTGTCCACCCCATAGGGTGTACCCACTGGTCTTGGAACACC
CATCCTTAATACGATGATTTTTCTGTCTGTGAAAATGAAGCCAGCAGGCTGCCCCTAGT
CAGTCCTTCCTTCCAGAGAAAAAGAGATTTGAGAAAGTGCCTGGGTAATTCACCATTAAT
TTCCCTCCCCCAAACCTCTCTGAGTCTTCCCTTAATATTTCTGGTGGTTCTGACCAAAGCAG
5 GTCATGGTTTGTGAGCATTGTTGGGATCCCAGTGAAGTAGATGTTTGTAGCCTTGCACTACT
TAGCCCTTCC

CF139275

CACGAGGTGGATTCCCCCGGCCTGGGTGGGGAGAGCGAGCTGGGTGCCCCCTAGATTCCC
10 CGCCCCCGCACCTCATGAGCCGACCCTCGGCTCCATGGAGCCCGGAATTATGCCACCTT
GGATGGAGCCAAGGATATCGAAGGCTTGCTGGGAGCGGGAGGGGGCGGAATCTGGTCGC
CCACTCCCCCTCTGAGCAGCCACCCAGCGGCGCCTACGCTGATGCCTGCTGTCAACTATGC
CCCCTTGGATCTGCCAGGCTCGGCGGAGCCGCCAAAGCAATGCCACCCATGCCCTGGGGT
GCCCCAGGGGACGTCCCCAGCTCCCGTGCCTTATGGTTACTTTGGAGGCGGGTACTACTC
15 CTGCCGAGTGTGCGGAGCTCGCTGAAACCCTGTGCCCAGGCA

CF122893

CACGAGGATTTTCTATCTAGAGCTCTGTAGAGCACTTTAGAAACCGCTTTCATGAATTGA
GCTAATTATGAATAAATTTGGAAGGCGATCCCTTTGCAGGGAAGCTTCTCTCAGACCCC
20 CTTCCATTACACCTCTCACCCCTGGTAACAGCAGGAAGACTGAGGAGAGGGGAACGGGCAG
ATTCGTTGTGTGGCTGTGATGTCCGTTTAGCATTTTTCTCAGCTGACAGCTGGGTAGGTG
GACAATTGTAGAGGCTGTCTCTTCCCTCCCTTGTCCACCCCATAGGGTGTACCCACTG
GTCTTGGAACACCCATCCTTAATACGATGATTTTTCTGTCTGTGAAAATGAAGCCAGC
AGGCTGCCCCCTAGTCAGTCCTTCCCTTCCAGAGAAAAAGAGATTTGAGAAAGTGCCTGGGT
25 AATTCACCATTAATTTCCCTCCCCCAAACCTCTCTGAGTCTTCCCTTAATATTTCTGGTGGT
TCTGACCAAAGCAGGTCATGGTTTGTGAGCATTGGGATCCCAGTGAAGTANATGTTTG
TAGCCTTGCACTTAGCCCTT

AI972423

CATTTTACACGACTGTAAAATCATCGTATTAAGGATGGGTGCTTCCAAGACCAGTGGGT
ACACCCTATGGGGTGGACAAGGAGGGAGGAAGAGACAGCCTCTACAATTGTCCACCTACC
CAGCTGTGAGCTGAGAAAAATGCTAAACGGACATCACAGCCACACAACGAATCTGCCCCGT
TCCCCCTCTCCTCAGTCTTCCCTGCTGTTACCAGGGTGAGAGGTGTAATGGAAGGGGGTCTG
AGAGAAAGCTTCCCTGCAAAGGGATCGCCTTCCAAATTTATTCATAATTAGCTCAATTCA
35 TGAAAGCGGTTTCTAAAGTGCTCTACAGAGCTCTAGATAGAAAATATGAGGCTAACGATC
ATGGCAGCTAGTACTGGTTATCGTGATTATTGCCACTGTCAGGATGAATGATTATGACTG
GGCCAGGTTCTTTGGGAACCCTGGTGGAGTGGGCTGTCACATG

AI918975

40 TGCAGCTAGTACTGGTTATCGTGATTATTGCCACTGTCAGGATGAATGATTATGACTGGG

CCAGGTTCTTTGGGAACCCTGGTGGAGTGGGCTGTCACATGGGGTTCCGTCTCCCTGCAC
ATACTGGGTACCCAGGCCGCTCCTGAGGAACAGTCCAGCACAGGGTTTCAGCGAGCTCCG
GGACACTCGGCCTCGTGC

5 AI826991

TTTTTTTTTTTTTTTTTTTTCTTTTTCACTTTCTCAAATCTCTTTTTCTCTGGAAGGAAG
GACTGACTAGGGGCAGCCTGCTGGCTTCATTTTACACCCACAAAAAATCATCGTATTAA
GGATGGGTGCTTCCAAAACAGTGGGTACACCCTATGGGGTGGACAAGGAGGGAGGAAAA
AACAGCCTCTACAATTGTCCACCTACCCAGCTGTCAGCTGAAAAAATGCTAAACGGACA
10 TCACAGCCACACAACGAATCTGCCCGTTCCCCTCTCCTCAGTCTTCCTGCTGTTACCAGG
GTGAAAGGTGTAATGGAAG

AI686312

ACCGACCCCACTACTTGCCACCGACCCGCTGCTCGGAGCTTCGGTTCTGCGGGTTGTCCA
15 GACTTCAGGCCTGTGCGCTCAATCGTGGAGAATGCGCCGGCAGGCCCCCCACCCCAAGCC
TAAGGTGCAGGAAGGACCAGCACGAACCCGCTGGCTTTGCTGCGCGGCCAGGAGATGAGT
CCCACCGGGCACTGAGCCCAGGTACAGGACATCAGAGAATGAACACAGAGGCAGAGGCC
TCATGTCCCTCTCAGAGTCCCGGCTCTGCAAAGAGCCCGTCTGTCTCCAGCTTCCAGAAT
TCCGCACTGTGAATCTGTCTACGTGGACTGGGAAAACAGGGTTGGCACCCTCTGCCACT
20 CCGTTTGTGCCTGGGAAGGGCTAAGTATGCAAGGCTACAAACATCTACTTCACTGGGATC
C

AI655923

TTTTTTTTTTTTTTTTCCCTGCAAAGGGATCGCCTTCCAAATTTATTCATAATTAGCTCAA
25 TTCATGAAAGCGGTTTCTAAAGTGCTCTACAGAGCTCTAGATAGAAAATATGAGGCTAAC
GATCATGGCAGCTAGTACTGGTTATCGTGATTATTGCCACTGTCAGGATGAATGATTATG
ACTGGGCCAGGTTCTTTGGGAACCCTGGTGGAGTGGGCTGTCACATGGGGTTCCGTCTCC
CTGCACATACTGGGTACCCAGGCCGCTCCTGA

30 CF146922

CACGAGGCGACTTGCGAGCTGGGAGCGATTTAAAACGCTTTGGATTCCCCGGCCTGGGTG
GGGAGAGCGAGCTGGGTGCCCCCTAGATTCCCCGCCCCCGCACCTCATGAGCCGACCCTC
GGCTCCATGGAGCCCGGCAATTATGCCACCTTGGATGGAGCCAAGGATATCGAAGGCTTG
CTGGGAGCGGGAGGGGGGCGGAATCTGGTCGCCCACTCCCCTCTGACCAGCCACCCAGCG
35 GCGCCTACGCTGATGCCTGCTGTCAACTATGCCCCCTTGGATCTGCCAGGCTCGGCGGAG
CCGCCAAAGCAATGCCACCCATGCCCTGGGGTGCCCCAGGGGACGTCCCCAGCTCCCGTG
CCTTATGGTTACTTTGGAGGCGGGTACTACTCCTGCCGAGTGTCCCGGAGCTCGCTGAAA
CCCTGTGCCCAGGCAGCCACCCTGGCCGCGTACCCCGCGGAGACTCCACGGCCGGGGAA
GAGTACCCCAAGCCGCCCCACTGAGTTTGCCTTCTATCCGGGATATCCGGGAACCTACCAG
40 CCTATGGCCAGTTACCTGGACGTGTCTGTGGTGCAGACTCTGGGTGCTCCTGGAGAACGC

GACATGACTCCCTGTTGCCTGTGGACAGTTACCAGTCTTGGGCTCTCGCTGGTGGCTGGA
ACAGCCAGATGTGTTGCCA

BF476369

5 GCGGCCGCGGCCCCACCACCAACTGCTCGCCATTTCGACCCCACTACTCGCCACCGACCCGC
TGCTCGGAGCTTCGGTTCTGCGGGTTGTCCAGACTTCAGGCCTGTGCGCTCAATCGTGGA
GAATGCGCCGGCAGGCCCCCACCACCCAGCCTAAGGTGCAGGAAGGACCAGCACGAACCC
GCTGGCTTTGCTGCGCGGCCAGGAGATGAGTCCCACCGGGCACTGAGCCCAGGTACAGGA
CATCAGAGAATGAACACAGAGGCAGAGGCCCTCATGTCCCTCTCAGAGTCCCGGCTCTGC
10 AAAGAGCCCGTCTGTCTCCAGCTTCCAGAATTCCGCACTGTGAATCTGTCTACGTGGACT
GGGAAAACAGGGTTGGCACCCTCTGCCACTCC

BF057410

15 GCGGCCGCGGCCCCACCACCAACTGCTCGCCACCGACCCCACTACTCGCCACCGACCCGCT
GCTCGGAGCTTCGGTTCTGCGGGTTGTCCAGACTTCAGGCCTGTGCGCTCAATCGTGGA
AATGCGCCGGCAGGCCCCCACCACCCAGCCTAAGGTGCAGGAAGGACCAGCACGAACCCG
CTGGCTTTGCTGCGCGGCCAGGAGATGAGTCCCACCGGGCACTGAGCCCAGGTACAGGAC
ATCAGAGAATGAACACAGAGGCAGAGGCCCTCATGTCCCTCTCAGAGTCCCGGCTCTGCA
AAGAGCCCGTCTGTCTCCAGCTTCCAGAATTCCGCACTGTGAATCTGTCTACGTGGACTG
20 GGAAAACAGGGTTGGCACCCTCTGCCACTCCGTTTGTGCCTGGGAAGGGCTAAGTATGC
AAGGCTACAAACATCTACTTCACTGGGATCCCAAATGCTCAACAAACCATGACCTGCTNT
GGTCAGAACCACCAGAAATATTAA

BE645544

25 GCGGCCGCGGCCCCACCACCAACTGCTCGCCACCGACCCCACTACTCGCCACCGACCCGCT
GCTCGGAGCTTCGGTTCTGCGGGTTGTCCAGACTTCAGGCCTGTGCGCTCAATCGTGGA
AATGCGCCGGCAGGCCCCCACCACCCAGCCTAAGGTGCAGGAAGGACCAGCACGAACCCG
CTGGCTTTGCTGCGCGGCCAGGAGATGAGTCCCACCGGGCACTGAGCCCAGGTACAGGAC
ATCAGAGAATGAACACAGAGGCAGAGGCCCTCATGTCCCTCTCAGAGTCCCGGCTCTGCA
30 AAGAGCCCGTCTGTCTCCAGCTTCCAGAATTCCGCACTGTGAATCTGTCTACGTGGACTG
GGAAAACAGGGTTGGCACCCTCTGCCACTCCGTTTGTGCCTGGGAAGGGCTAAGTATGC
AAGGCTACAAACATCTACTTCACTGGGATCC

BE645408

35 TCCTCCCTCTAAGAAAGGCGCAAGCGTCAAGAGGGTGCTGCCCGCTGGTTTCTGCAAATG
CTGCCTTCCAAAAGGACCTGGTGGGTTCTGTTCTCCCTGGCAACACATCTGGCTGTTCC
AGCCACCAGCGAGAGCCCAAGACTGGTAACTGTCCACAGGCAACAGGGAGTCATGTGCGC
GTTCTCCAGGAGCACCCAGAGTCTGCACCACAGACACGT

40 **BE388501**

TTAATACGATGATTTTTCTGTCGTGTGAAAATGAAGCCAGCAGGCTGCCCCCTAGTCAGTC
CTTCCTTCCAGAGAAAAAGAGATTTGAGAAAGTGCCTGGGTAATTCACCATTAATTTCCCT
CCCCCAAACCTCTCTGAGTCTTCCCTTAATATTTCTGGTGGTTCTGACCAAAGCAGGTCAT
GGTTTGGTTGAGCATTGTTGGGATCCCAGTGAAGTAGATGTTTGTAGCCTTGCATACTTAGCC
5 CTTCCCAGGCACAAACGGAGTGGCAGAGTGGTGCCAACCCTGTTTTCCCAGTCCACGTAG
ACAGATTCACAGTGCAGGAATCTGGAAGCTGGAGACAGACGGGCTCTTTGCAGAGCCGGG
ACTCTGAGAGGGACATGAGGGCCTCTGCCTCTGTGTTTCACTCTCTGATGTCCTGTACCTG
GGCTCAGTGCCCGGTGGGACTCATCTCCTGGCCGCGCAGCAAAGCCAGCGGGTTCGTGCT
GGTCCTTCTCTGCACCTTAGGCTGGGGGTGGGGGGCCTGCCGGCGCATTCTCCACGATTGA
10 GCGCACAGGCCTGAAGTCTGGACAACCCGCGAGAACCAGCTCCGAGCAGCGGGTCGGTG
GCGAGTAGTGGGGGTCGGTGGCGAACAAGTGGTGGTGGGCCGGGGCCGCATAACTCGAGG
ACTTTCCTCCCGGAGCAGTCCCTAAAAACCCGGGGGCGC

CF147366

15 GACGAGGACAATTGTAGAGGCTGTCTCTTCCTCCCTCCTTGTACCCCATAGGGTGTACC
ACTGGTCTTGGAAGCACCCATCCTTAATACGATGATTTTTCTGTCGTGTGAAAATGAAGC
CAGCAGGCTGCCCCCTAGTCAGTCCTTCCCTTCCAGAGAAAAAGAGATTTGAGAAAGTGCCT
GGGTAATTCACCATTAATTTCCCTCCCCCAAACCTCTCTGAGTCTTCCCTTAATATTTCTGG
TGGTCTGACCAAAGCAGGTCATGGTTTGTGAGCATTGTTGGGATCCCAGTGAAGTAGATG
20 TTTGTAGCCTTGCATACTTAGCCCTTCCCAGGCACAAACGGAGTGGCAGAGTGGTGCCAA
CCCTGTTTTCCCAGTCCACGTAGACAGATTCACAGTGCAGGAATCTGGAAGCTGGAGACA
GACGGGCTCTTTGCAGAGCCGGGACTCTGAGAGGGACATGAGGGCCTCTGCCTCTGTGTT
CATTCTCTGATGTCCTGTACCTGGGCTCAGTGCCCGGTGGGACTCATCTCCTGGCCGCGC
AGCAAAGCCAGCGGGTTCGTGCTGGTCCTTCCTGC

25

CF147143

CACGAGGCGACTTGCGAGCTGGGAGCGATTTAAAACGCTTTGGATTCCCCCGGCCTGGGT
GGGGAGAGCGAGCTGGGTGCCCCCTAGATCCCCGCCCCCGCACCTCATGAGCCGACCCT
CGGCTCCATGGAGCCCGCAATTATGCCACCTTGATGGAGCCAAGGATATCGAAGGCTT
30 GCTGGGAGCGGGAGGGGGGCGGAATCTGGTCGCCCACTCCCCTCTGACCAGCCACCCAGC
GGCGCCTACGCTGATGCCTGCTGTCAACTATGCCCCCTTGATCTGCCAGGCTCGGCGGA
GCCGCCAAAGCAATGCCACCCATGCCCTGGGGTGCCCCAGGGGACGTCCCCAGCTCCCGT
GCCTTATGGTTACTTTGGAGGCGGGTACTACTCCTGCCGAGTGTCCCGGAGCTCGTGAA
ACCCTGTGCCAGGCAGCCACCCTGGCCGCGTACCCCGCGGAGACTCCCACGGCCGGGGA
35 AGAGTACCCAGCCGCCCCACTGAGTTTGCCTTCTATCCGGGATATCCGGGAACCTACAG
CCTATGGCCAGTTACCTGGACGTGTCTGTGGTGCAGACTCTGGGTGCTCCTGGAGAACGC
GACATGACTCCCTGTTGCCCTGTGGACAGTTACCAATCTTGGGCTCTCGCTGGTGGCTGGA
ACAGCCAGATGTGTTGCCAGGGAG

40 BT007410

atggagcccc gcaattatgc caccttggat ggagccaagg atatcgaagg cttgctggga

gcgggagggg ggcggaatct ggtcgccac tcccctctga ccagccaccc agcggcgcct
acgctgatgc ctgctgtcaa ctatgcccc ttggatctgc caggctcggc ggagccgcca
aagcaatgcc acccatgccc tggggtgccc caggggacgt cccagctcc cgtgccttat
ggttactttg gaggcgggta ctactcctgc cgagtgtccc ggagctcgct gaaaccctgt
5 gcccaggcag ccaccctggc cgcgtacccc gcgagactc ccacggccgg ggaagagtac
cccagccgcc ccactgagtt tgccttctat ccgggatatc cgggaaccta ccagcctatg
gccagttacc tggacgtgtc tgtggtgcag actctgggtg ctcttgga accgcgacat
gactccctgt tgcctgtgga cagttaccag tcttgggctc tcgctgggtg ctggaacagc
cagatgtgtt gccagggaga acagaaccca ccaggtccct tttggaaggc agcatttgca
10 gactccagcg ggcagcacc tcttgacgcc tgcgcctttc gtcgcggccg caagaaacgc
attccgtaca gcaaggggca gttgcgggag ctggagcggg agtatgcggc taacaagttc
atcaccaagg acaagaggcg caagatctcg gcagccacca gcctctcgga gcgccagatt
accatctggt ttcagaaccg ccgggtcaaa gagaagaagg ttctcgccaa ggtgaagaac
agcgctaccc cttag

15

BC007092

ggattccccc ggcctgggtg gggagagcga gctgggtgcc ccctagattc cccgcccccg
cacctcatga gccgaccctc ggctccatgg agcccgcaa ttatgccacc ttggatggag
20 ccaaggatat cgaaggcttg ctgggagcgg gaggggggcg gaatctggtc gccactccc
ctctgaccag ccaccagcg gcgcctacgc tgatgcctgc tgtcaactat gcccccttgg
atctgccagg ctcggcggag ccgccaaagc aatgccaccc atgccctggg gtgccccagg
ggacgtcccc agctcccgtg ccttatggtt actttggagg cgggtactac tcctgccgag
tgtcccggag ctcgctgaaa ccctgtgccc aggcagccac cctggcccg tgagtttgcc ttctatccgg
agactccac ggccggggaa gagtaccca gccgccccac tgagtttgcc ttctatccgg
25 gatatccggg aacctaccag cctatggcca gttacctgga cgtgtctgtg gtgcagactc
tgggtgctcc tggagaaccg cgacatgact cctggttgcc tgtggacagt taccagtctt
gggctctcgc tgggtggctgg aacagccaga tgtgttgcca gggagaacag aaccaccag
gtcccttttg gaaggcagca tttgcagact ccagcgggca gcaccctcct gacgcctgcg
cctttcgtcg cggccgcaag aaacgcattc cgtacagcaa ggggcagttg cgggagctgg
30 agcgggagta tgcggctaac aagttcatca ccaaggacaa gaggcgcaag atctcggcag
ccaccagcct ctcgagcgc cagattacca tctggtttca gaaccgccgg gtcaaagaga
agaaggttct cgccaagggtg aagaacagcg ctaccctta agagatctcc ttgcctgggt
gggaggagcg aaagtggggg tgtcctgggg agaccaggaa cctgccaagc ccaggctggg
gccaaggact ctgctgagag gcccctagag acaacaccct tcccaggcca ctggctgctg
35 gactgttcct caggagcggc ctgggtaccc agtatgtgca gggagacgga accccatgtg
acagcccact ccaccagggt tcccaaagaa cctggcccag tcataatcat tcactctgac
agtggcaata atcacgataa ccagtactag ctgccatgat cgttagcctc atattttcta
tctagagctc tgtagagcac tttagaaacc gctttcatga attgagctaa ttatgaataa
at ttg gaaaa aaaaaaaaa aaaaaaaa aaaaaa

40

U57052

cgggtgcccc ctagattccc cgcccccgca cctcatgagc cgaccctcgg ctccatggag

	cccggcaatt	atgccacctt	ggatggagcc	aaggatatcg	aaggcttgct	gggagcggga
	ggggggcgga	atctggtcgc	ccactcccct	ctgaccagcc	acccagcggc	gcctacgctg
	atgcctgctg	tcaactatgc	ccccttggat	ctgccaggct	cggcggagcc	gccaaagcaa
	tgccacccat	gccctggggt	gccccagggg	acgtcccag	ctcccgtgcc	ttatggttac
5	tttggaggcg	ggtactactc	ctgccgagtg	tcccggagct	cgtgaaacc	ctgtgcccag
	gcagccaccc	tggccgcgta	ccccgcggag	actcccacgg	cggggaaga	gtaccccagc
	cgccccactg	agtttgcctt	ctatccggga	tatccgggaa	cctaccacgc	tatggccagt
	tacctggacg	tgtctgtggt	gcagactctg	ggtgctcctg	gagaaccgcg	acatgactcc
	ctgttgccctg	tggacagtta	ccagtcttgg	gctctcgctg	gtggctggaa	cagccagatg
10	tgttgccagg	gagaacagaa	cccaccaggt	cccttttggg	aggcagcatt	tgcagactcc
	agcgggcagc	accctcctga	cgctcccgcc	tttcgtcgcg	gccgaagaa	acgcattccg
	tacagcaagg	ggcagttgcg	ggagctggag	cgggagtatg	cggctaacaa	gttcatcacc
	aaggacaaga	ggcgcaagat	ctcggcagcc	accagcctct	cggagcgcca	gattaccatc
	tggtttcaga	accgccgggt	caaagagaag	aaggttctcg	ccaaggtgaa	gaacagcgct
15	acccttaag	agatctcctt	gcctgggtgg	gaggagcgaa	agtgggggtg	tcctggggag
	accaggaacc	tgccaagccc	aggctggggc	caaggactct	gctgagaggc	ccctagagac
	aacacc					

U81599

20	tcctaatacg	actcactata	gggctcgagc	ggccgcccgg	gcaggtcgaa	tgcaggcgac
	ttgcgagctg	ggagcgattt	aaaacgcttt	ggattccccc	ggcctgggtg	gggagagcga
	gctgggtgcc	ccctagattc	cccgcccccg	cacctcatga	gccgaccctc	ggctccatgg
	agcccgga	ttatgccacc	ttggatggag	ccaaggatat	cgaaggcttg	ctgggagcgg
	gagggggggcg	gaatctggtc	gcccactccc	ctctgaccag	ccaccagcgc	gcgccctacgc
25	tgatgcctgc	tgtcaactat	gcccccttgg	atctgccagg	ctcggcggag	ccgccaagc
	aatgccaccc	atgccttggg	gtgccccagg	ggacgtcccc	agctcccgtg	ccttatgggtt
	actttggagg	cgggtactac	tcctgccgag	tgtcccggag	ctcgctgaaa	ccctgtgccc
	aggcagccac	cctggccgcg	taccccgcg	agactccac	ggccggggaa	gagtacccca
	gtcgccccac	tgagtttgcc	ttctatccgg	gatatccggg	aacctaccac	gctatggcca
30	gttacctgga	cgtgtctgtg	gtgcagactc	tgggtgctcc	tggagaaccg	cgacatgact
	ccctgttgcc	tgtggacagt	taccagtctt	gggctctcgc	tgggtggctgg	aacagccaga
	tgtgttgcca	gggagaacag	aaccaccag	gtcccttttg	gaaggcagca	tttgagact
	ccagcgggca	gcaccctcct	gacgcctgcg	cctttcgtcg	cggccgcaag	aaacgcattc
	cgtacagcaa	ggggcagttg	cgggagctgg	agcgggagta	tgcggtaac	aagttcatca
35	ccaaggacaa	gaggcgcaag	atctcggcag	ccaccagcct	ctcggagcgc	cagattacca
	tctggtttca	gaaccgcggg	gtcaaagaga	agaaggttct	cgccaagggtg	aagaacagcg
	ctacccttta	agagatctcc	ttgcctgggt	gggaggagcg	aaagtggggg	tgtcctgggg
	agaccagaaa	cctgccaaagc	ccaggctggg	gccaaggact	ctgctgagag	gcccctagag
	acaacaccct	tcccaggcca	ctggctgctg	gactgttctt	caggagcggc	ctgggtaccc
40	agtatgtgca	gggagacgga	accccatgtg	acaggcccac	tccaccaggg	ttcccaaaga
	acctggccca	gtcataatca	ttcatcctca	cagtggcaat	aatcacgata	accagt

CB120119

ATTTTCTGTCGTGTGAAAATGAAGCCAGCAGGCTGCCCCTAGTCAGTCCTTCCTTCCAG
AGAAAAAGAGATTTGAGAAAGTGCCTGGGTAATTCACCATTAATTCCTCCCCCAAACCTC
TCTGAGTCTTCCCTTAATATTTCTGGTGGTTCTGACCAAAGCAGGTCATGGTTTGTGAG
CATTGGGATCCCACTGAAGTAGATGTTTGTAGCCTTGCATACTTAGCCCTTCCCAGGCA
5 CAAACGGAGTGGCAGAGTGGTGCCAACCTGTTTTCCAGTCCACGTAGACAGATTCACA
GTGCGGAATTCTGGAAGCTGGAGACAGACGGGCTCTTTGCAGAGCCGGGACTCTGAGAGG
GACATGAGGGCCTCTGCCTCTGTGTTTCTCTGATGTCCTGTACCTGGGCTCAGTGCC
CGGTGGGACTCATCTCCTGGCTGCGCAGCAAAGCCAGCGGGTTCGTGCTGGTCCTTCCTG
CACCTTAGGCTGGGGGTGGGGGGCCT

CB125764

ATTTTCTGTCGTGTGAAAATGAAGCCAGCAGGCTGCCCCTAGTCAGTCCTTCCTTCCAG
AGAAAAAGAGATTTGAGAAAGTGCCTGGGTAATTCACCATTAATTCCTCCCCCAAACCTC
TCTGAGTCTTCCCTTAATATTTCTGGTGGTTCTGACCAAAGCAGGTCATGGTTTGTGAG
15 CATTGGGATCCCACTGAAGTAGATGTTTGTAGCCTTGCATACTTAGCCCTTCCCAGGCA
CAAACGGAGTGGCAGAGTGGTGCCAACCTGTTTTCCAGTCCACGTAGACAGATTCACA
GTGCGGAATTCTGGAAGCTGGAGACAGACGGGCTCTTTGCAGAGCCGGGACTCTGAGAGG
GACATGAGGGCCTCTGCCTCTGTGTTTCTCTGATGTCCTGTACCTGGGCTCAGTGCC
CGGTGGGACTCATCTCCTGGCTGCGCAGCAAAGCCAGCGGGTTCGTGCTGGTCCTTCCTG
20 CACCTTAGGCTGGGGGTGGGGGGGCCTGCCGGCGCATTCTCCACGATTGAGCGCACAGG
CCTGAAGTCTGGACAACCCGCAGAACCGAAGCTCCGAGCAGCGGGTTCGGTGGCGAGT

AU098628

ATTTAAACGCTTTGGATTCTTTCGTCTGCGTGGGAGAGCGAGCTGGGTGCCCCCTAG
25 ATTCCCCGCCCCCGCACCTCATGAGCCGACCCTCGGCTCCATGGAGCCCGGCACTTATGC
CACCTTGGATGGAGCCAAGGATATCGAAGGCTTGCTGGGAGCGGGAGGGGGCGGAATCT
GGTCGCCCCACTCCCCCTCTGACCAGCCACCCAGCGGCGCTACGCTGATGCCTGCTGTCAA
TTATGCCCCCTTGCATCTGCCAGGCTCGGCGGAGCCGCCAAAGCAATGCCACCCATGCCC

CB126130

ATTTTCTGTCGTGTGAAAATGAAGCCAGCAGGCTGCCCCTAGTCAGTCCTTCCTTCCAG
AGAAAAAGAGATTTGAGAAAGTGCCTGGGTAATTCACCATTAATTCCTCCCCCAAACCTC
TCTGAGTCTTCCCTTAATATTTCTGGTGGTTCTGACCAAAGCAGGTCATGGTTTGTGAG
CATTGGGATCCCACTGAAGTAGATGTTTGTAGCCTTGCATACTTAGCCCTTCCCAGGCA
35 CAAACGGAGTGGCAGAGTGGTGCCAACCTGTTTTCCAGTCCACGTAGACAGATTCACA
GTGCGGAATTCTGGAAGCTGGAGACAGACGGGCTCTTTGCAGAGCCGGGACTCTGAGAGG
GACATGAGGGCCTCTGCCTCTGTGTTTCTCTGATGTCCTGTACCTGGGCTCAGTGCC
CGGTGGGACTCATCTCCTGGCTGCGCAGCAAAGCCAGCGGGTTCGTGCTGGTCCTTCCTG
CACCTTAGGCTGGGGGTGGGGGGCCTGC

BI023924

AGGCCGCACCCAGTCTTAAGGTGCAGTGAAGGACAGCACGAACCCGCTGTGCTTTGCTGC
GCGGCAGGAGATGAGTCCCACCGGGCACTGAGCCAGGTACAGGACATCAGAGAATGAAC
ACAGAGGCAGAGGCCCTCATGTCCCTCTCAGAGTCCCGGCTCTGCAAAGAGCCCGTCTGT
CTCCAGCTTCCAGAATTCCGCACTGTGAATCTGTCTACGTGGACTGNGAAAACAGGGTTG
5 GCACCACTCTGCCACTCCGTTTGTGCCTNGGGGCGGGCAGAGGG

BM767063.1

AAAAACGCTTTGGATTCCCCCGGCCTGGGTGGGGAGAGCGAGCTGGGTGCCCCCTAGATT
CCCCGCCCCCGCACCTCATGAGCCGACCTCGGCTCCATGGAGCCCGGCAATTATGCCAC
10 CTTGGATGGAGCCAAGGATATCGAAGGCTTGCTGGGAGCGGGAGGGGGCGGAATCTGGT
CGCCCACTCCCCTCTGACCAGCCACCCAGCGGCGCCTACGCTGATGCCTGCTGTCAACTA
TGCCCCCTTGGATCTGCCAGGCTCGGCGGAGCCGCCAAAGCAATGCCACCCATGCCCTGG
GGTGCCCCAGGGGACGTCCCCAGCTCCCGTGCCCTTATGGTTACTTTGGAGGCGGGTACTA
CTCCTGCCGAGTGTCCCGGAGCTCGCTGAAACCCTGTGCCAGGCAGCCACCCTGGCCGC
15 GTACCCCGCGGAGACTCCCACGGCCGGGGAAGAGTACCCAGCCGCCCCACTGAGTTTGC
CTTCTATCCGGGATATCCGGGAACCTACCAGCCTATGGCCAGTTACCTGGACGTGTCTGT
GGTGCAGACTCTGGGTGCTCCTGGAGAACC GCGACATGACTCCCTGTTGCCTGTGGACAG
TTACCAGTCTTGGGCTCTCGCTGGTGGCTGGAACAGCCAGATGTGTTGCCA

BM794275

GCAGACTCTGGGTGCTCCTGGAGAACCGCGACGTGACTCCCTGTTGCCTGTGGACAGTTA
CCACTCTTGGGCTCTCGCTGGTGGCTGGAACAGCCAGATGTGTTGCCAGGGAGAACAGAA
CCCACCAGGTCCCTTTTGGAAAGGCAGCATTTGCAGACTCCAGCGGGCAGCACCCCTCCTGA
CGCCTGCGCCTTTTCGTGCGGCGCGCAAGAAACGCATTCCGTACAGCAAGGGGCAGTTGCG
25 GGAGCTGGAGCGGGAGTATGCGGCTAACAAGTTCATCACCAAGGACAAGAGGCGCAAGAT
CTCGGCAGCCACCAGCCTCTCGGAGCGCCAGATTACCATCTGGTTTCAGAACCGCCGGGT
CAAAGAGAAGAAGGTTCTCGCCAAGGTGAAGAACAGCGCTACCCCTTAAGAGATCTCCTT
GCCTGGGTGGGAGGATCTAAAGTGGGGGTGTCCTGGGGAGACCAGGAACCTGCCAAGCCC
AGGCTGGGGCCAAGGACT

BQ363211

ACGCTGCACTGCGTTTCAAAGAGAAGAAGGTTCTCGCCAAGGTGAAGAACAGCGCTACCC
CTTAAGAGATCTCCTTGCTTGGGTGGGAGGAGCGAAAGTGGGGGTGTCCTGGGGAGACCA
GGAACCTGCCATCACCAAGGCTGGGCCCAAGGACTCTGCTGAGAGGCCCTAGAGACAACA
35 CCCTTCCCAGGCCATTGCTTGCTGGACTGTGCCTCAGGAGCGGCCTGGGTACC

BM932052

GAGTTTTCCAATTTCCAAAGAAAAATTTAGGTTTCCTGCAGCCGTGACATATGTGTGTGC
ACTGGGATGGGTTAATGTGTGTGTGTGTGTGTATGCGCATGTATTGGGAGTGGGGGCA
40 GAAACGTGTTTCCAGAATTTGCCTGTAGAATCTAAAAGAGTGGCCAAGAGTCTGGAAATG

CATGAAGACTGGACGTATGTGATGGTGGGCAAAGGCCTGACTGTGTGTGGTGTGTGGGTA
TGTTTGCAGATTCGCGGGTGTGAGAGCAGTGATGGGTGAGGGTGGCCTTCAGGAGCCAAG
GCTGATCGGTGGTGAGAGAACAAGCCGGAAGCCAGGGTGCTGTCTGGTATGCTTTGGAG
GAACAGGATTGCACGTGCGCCTGTAGGGTGACCTGTGTGCACCTGTGAGATGACTTAGCT
5 TGGGGCTTGCAAGGCCTGGGTCTGCATGGGTGGGTATCTGACCATGCCTTTTCTCTCCCTC
CCTTTCACGCCGCGCAGACTCCAGCGGGCAGCACCTCCTGACGCCTGCGCCTTTCGTC

AA357646.1

10 CCGGCCTGGGTGGGGAGAGCGAGCTGGGTGCCCCCTAGATTCCCCGCCCCCGCACCTCAT
GAGCCGACCCTCGGCTCCATGGAGCCCGGCAATTATGCCACCTTGATGGAGCCAAGGAT
ATCGAAGGCTTGCTGGGAGCGGGAGGGGGGCGGAATCTGGTCGCCCCTCCCTCTGACC
AGCCACCCAGCGGCGCCTACGCTTGATGCCTGCTTGCTCAACTATGCCCCCTTGATCTGC

AW609525

15 ACCGCGGGTCAAATTTATTTCATAATTAGCTCAATCATGAAAGCGGTTCTAAAGTGCTCTA
CAGAGCTCTAGATAGAAAATATGAGGCTAACGATCATGGCAGCTAGTACTGGTTATCGTG
ATTATGGCCACTGTCAGGATGAATGATAATGACTGGGCCAGGTCCTTTGGAAACCCTGGT
GGAGTGGGCTGTCACATGGGGTCCCGTCTCCCTGCACATACTGGGTACCCAGGCCGCTCC
TGAGGAACAGTCCAGCAGCCAGTGGCCTGGGAAGGGTGTGGTCTCTAGGGGCCTCTCAGC
20 AGAGTCCTTGCCCCAGCCTGGGCTTGGCAGGTCCCTGGTCTCCCCAGGACACCCCCACT
TTCGCTCCTCCCACCCAGGCAAGGAGATCTCTTAAGGGGTAGCGCTGTTCTTCACCTTGG
CGAGAACCTTCTTCTTTGAACCGGCGGTGCGGCGTGGGGTACCGAGC

CB126919

25 ATTTTTCTGTCGTGTGAAAATGAAGCCAGCAGGCTGCCCCCTAGTCAGTCCTTCCTTCCAG
AGAAAAAGAGATTTGAGAAAGTGCCTGGGTAAATTCACCATTAATTTCTCTCCCCCAAACCTC
TCTGAGTCTTCCCTTAATATTTCTGGTGGTTCTGACCAAAGCAAGTCATGGTTTGTGAG
CATTTGGGATCCCAGTGAAGTAGATGTTTGTAGCCTTGCATACTTAGCCCTTCCCAGGCA
CAAACGGAGTGGCAGAGTGGTGCCAACCCTGTTTTCCCACTCCACGTAGACAGATTACACA
30 GTGCGGAATTCTGGAAGCTGGAGACAGACGGGCTCTTTGCAGAGCCGGGACTCTGAGAGG
GACATGAAGGCCTCTGCCTCTGTGTTTATTCTCTGATGTCCTGTACCTGGGCTCAGTGCC
CGGTGGGACTCATCTCCTGGCTGCGCAGCAAAGCCAGCGGGTTCTGTGCTGGT

AW609336

35 CCAACGAGAAGAAGGTTCTCGCAAGGTGAAGAACAGCGCTACCCCTTAAGAGATCTCCTT
GCGTGGGTGGGAGGAGCGAAAGTGGGGGTGTCCTGGGGAGACCAGGAACCTGCCAGCCCCA
GGCTGAGGCCAAGGACTCTGCTGAGAGGCCCCCTAGAGACAACACCCTTCCCAGGCCACTG
GATGCTGAACTGTCCCTCAGGAGCGGCCTGGGTACCCAGTATGTGCAGGGAGACGGAACC
CCATGTGACAGCCCCTCCACCAGGGTTCCCAAAGAACCTGGCCCCAGTCATAATCATTC
40 ATCCTGACAGTGGCAATAATCACGATAACCAGTACTAGCTGCCATGATCGTAAGCCTCAT

ATTTGCTATCTAGAGCTCTGTAGAGCACTTTAGAAACCGCTTTCATGAATTGAGCTAATT
ATGACTCAATTTGAACCGGCGTCCGGCGTG

AW609244

5 ACGCGCACCGCGGTCAAGAGAAGAAGGTTCTCGCAAGGTGAAGAACAGCGCTACCCCTTA
AGAGATCTCCTTGCGTGGGTGGGAGGAGCGAAAGTGGGGTGTCTGGGGAGACCAGGAA
CCTGCCAAGCCCAGGCTGTGGCCAAGGACTCTGCTGAGAGGCCCTATGAGACAACACCC
TTCCCAGGCCACTGGCTGCTGGGACTGTTCCCTCAGGAGCGGCCTGGGTACCCGAGTAATG
TGCAGGGGAGACGGAACCCCATGTGACAGCCCACTCCACCAGGGTTCCCAAAGAACCCT
10 GGCCCAGTCATAATCATTTCATCCTGACAGTGGCAATAATCACGATAACCAGTACTAGCTG
CCATGATCGTAAGCCTCATATTTGCTATCTAGAGCTCTGTAGAGCCCTTTAGAAACCGCT
TTCATGAATGGAGCTAAATTATGAATACATTTGAACCGGCGATCCGACGTGA

BF855145

15 CTAGAGGATCCCGGAAGCAACTGCAACAGGTTCCCAAAGAACCGGGCCAGTCATAATCAT
TCATCCTGACAGGGCAATAATCACGATAACCAGTACTAGCTGCCATGATCGTTAGCCTCA
TATTTTCTATCTAGAGCTCTGTAGAGCACTTTAGAAACCGCTTTCATGAATGGAGCTAAT
TATGAATAAATTTGGAAGGCGATCCCTTGGCAGGGAAGCTTTCTCTCAGACCCCTTCCA
TTACACCTCTCACCTGGTAACAGCAGGAAGACTGAGGAGAGGGGAACGGGCAGATTCTGT
20 GGTGTTGCAGTGTGCTTCCG

AU126914

GAGCGAATGCAGGCGACTTGCGAGCTGGGAGCGATTTAAACGCTTTGGATTCCCCCGGC
CTGGGTGGGGAGAGCGAGCTGGGTGCCCCCTAGATTCCCCGCCCCCGCACCTCATGAGCC
25 GACCCTCGGCTCCATGGAGCCCCGGCAATTATGCCACCTTGGATGGAGCCAAGGATATCGA
AGACTTGCTGGGAGCGGGAGGGGGGCGGAATCTGGTGCGCCCACTCCCCTCTGACCAGCCA
CCCAGCGGCGCCTACGCTGATGCCTGCTGTCAACTATGCCCCCTTGGATCTGCCAGGCTC
GGCGGAGCCGCCAAAGCAATGCCACCCATGCCCTGGGGTGCCCCAGGGGACGTCCCCAGC
TCCCGTGCCCTTATGGTTACTTTGGAGGCGGGTNCTACTCCTGCCGAGTGTCCCGGAGCTC
30 GCTGAAACCCTGTGCCCCANNCANCCACCCTGGCCGCGTN

CB126449

CTCTGCCTCTGTGTTTCAATCTCTGATGTCCTGTACCTGTGCTCAGTGCCCGGTGGGACTC
ATCTCCTGGCTGCGCAGCAAAGCCAGCGGGTTCGTGCTGGTCCTTCCTGCACCTTCGGCT
35 GGGGGTGGGGGGCCTGCCGGCGCATCTCCACGATT

AW582404

ACGCTGCACCGCCGGTCCAAGAGAAGAAGGTTCTCGCCAAGGTGAAGAACAGCGCTACCC
CTTTAAGAGATCTCCTTGCTGGGGTGGGAGGAGCGAAAGTGGGGTGTCTGGGGAGACCA

GGAACCTGCCAGCCCCAGGCTGGGCCCCAAGGACTCTGCTGAGAGGCCCTAGAGACAACA
CCCTTCCCAGGCCACTGTCTGCTGGACTGTTCTCAGGAGCGGCCTGGGTACNCAGTATG
TGCAGGGAGACGGAACCCCATGTGACAGCCCACTCCACCAGGGTCCCAAAGAACCTGGC
CCAGTCATAATCATTCATCCTGACAGTGGCAATAATCACGATAACCAGTACTAGCTGCCA
5 TGATCGTTAGCCTCATATTTTCTATCTAGAGCTCTGTAGAGCACTTTAGAAACCGCTTTC
ATGAATTGAGCTACTTATGAATCACTTTGAACCGGCGGTGCGGCGTG

BX641644

GGGGGAGAGCGAGCTGGGTGCCCCCTAGATTCCCCGCCCCCGCACCTCATGAGCCGACCC
10 TCGGCTCCATGGAGCCCCGGCAATTATGCCACCTTGATGGAGCCAAGGATATCGAAGGCT
TGCTGGGAGCGGGAGGGGGGCGGAATCTGGTCGCCCCACTCCCCTCTGACCAGCCACCCAG
CGGCGCCTACGCTGACGCTGCTGTCAACTATGCCCCCTTGATCTGCCAGGCTCGGCGG
AGCCGCCAAAGCAATGCCACCCATGCCCTGGGGTGCCCCAGGGGACGTCCCCAGCTCCCCG
TGCTTATGGTTACTTTGGAGGCGGGTACTACTCCTGCCGAGTGTCCCGGAGCTCGCTGA
15 AACCTGTGCCAGGCAGCCACCTGGCCGCGTACCCCGCGGAGACTCCCACGGCCGGGG
AAGAGTACCCAGCCGCCCACTGAGTTTGCCTTCTATCCGGGATATCCGGGAACCTACC
AGCCTATGGCCAGTTACCTGGACGTGTCTGTGGTGACAGTCTGGGTGCTCCTGGAGAAC
CGCGACATGACTCCCTGTTGCCTGTGGACAGTTACCAGTCTTGGGCTCTCGCTNGTGGCT
GGAACAGCCAGATGTGTTGCCAGGGAGAAACAGAACCCACCAGGTCCCTTTTGAAGGCAG
20 CATTTG

Sequences from Table 4 not disclosed above

AW006861 (IMAGE Clone ID: :2497262)

GCTGAGTTCTGAAGCTTCTGAGTTCTGCAGCCTCACCTCTGAGAAAACCTCTTTTCCACC
AATACCATGAAGCTCTGCGTGACTGTCTCTCTCCTCATGCTAGTAGCTGCCTTCTGC
TCTCTAGCGCTCTCAGCACCAATGGGCTCAGACCCTCCCACCGCCTGCTGCTTTTCTTAC
ACCGCGAGGAAGCTTCCTCGCAACTTTGTGGTAGATTACTATGAGACCAGCAGCCTCTGC
TCCCAGCCAGCTGTGGTATTCCAAACCAAAAGAAGCAAGCAAGTCTGTGCTGATCCCAGT
30 GAATCCTGGGTCCAGGAGTACGTGTATGACCTGGAAGTGAAGTGAAGTCTGCTCAGAGACAG
GAAGTCTTCAGGGAAGGTCACCTGAGCCCGGATGCTTCTCCATGAGACACATCTCCTCCA
TACTCAGGACTCCTCTCCGCAGTTCCTGTCCCTTCTCTTAATTTAATCTTTTTTATGTGC
CGTGTTATTGTATTAGGTGTCATTTCCATTATTTATATTAGTTTAGCCAAAGGATAAGTG
TCCCCTATGGGGATGGTCCACTGTCACTGTTTCTCTGCTGTTGCAAATACATGGATAACA
35 CATTTGATTCTGTGTGTTTTTATAATAAACTTTAAATAAAATGCAAAAAAAAAAAAAA
AAAA

X59770

GCCACGTGCTGCTGGGTCTCAGTCCTCCACTTCCCGTGTCTCTGGAAGTTGTCAGGAGC
40 AATGTTGCGCTTGTACGTGTTGGTAATGGGAGTTTCTGCCTTCACCCTTCAGCCTGCGGC

ACACACAGGGGCTGCCAGAAGCTGCCGGTTTTCGTGGGAGGCATTACAAGCGGGAGTTTCAG
GCTGGAAGGGGAGCCTGTAGCCCTGAGGTGCCCCAGGTGCCCTACTGGTTGTGGGCCTC
TGTCAGCCCCCGCATCAACCTGACATGGCATAAAAAATGACTCTGCTAGGACGGTCCCAGG
AGAAGAAGAGACACGGATGTGGGCCCAGGACGGTGCTCTGTGGCTTCTGCCAGCCTTGCA
5 GGAGGACTCTGGCACCTACGTCTGCACTACTAGAAATGCTTCTTACTGTGACAAAATGTC
CATTGAGCTCAGAGTTTTTGAGAATACAGATGCTTTCCTGCCGTTTCATCTCATACCCGCA
AATTTTAACCTTGTC AACCTCTGGGGTATTAGTATGCCCTGACCTGAGTGAATTCACCCG
TGACAAAACCTGACGTGAAGATTCAATGGTACAAGGATTCTCTTCTTTTGGATAAAGACAA
TGAGAAATTTCTAAGTGTGAGGGGGACCACTCACTTACTCGTACACGATGTGGCCCTGGA
10 AGATGCTGGCTATTACCGCTGTGTCTGACATTTGCCCATGAAGGCCAGCAATACAACAT
CACTAGGAGTATTGAGCTACGCATCAAGAAAAAAAAAGAAGAGACCATTCTGTGATCAT
TTCCCCCCTCAAGACCATATCAGCTTCTCTGGGGTCAAGACTGACAATCCCGTGTAAGGT
GTTTCTGGGAACCGGCACACCCTTAACCACCATGCTGTGGTGGACGGCCAATGACACCCA
CATAGAGAGCGCCTACCCGGGAGGCCGCGTGACCGAGGGGCCACGCCAGGAATATTCAGA
15 AAATAATGAGA ACTACATTGAAGTGCCATTGATTTTTTGATCCTGTCAAGAGAGGATTT
GCACATGGATTTTTAAATGTGTTGTCCATAATACCCTGAGTTTTTCAGACACTACGCACCAC
AGTCAAGGAAGCCTCCTCCACGTTCTCCTGGGGCATTGTGCTGGCCCCACTTTCCTGCTGGC
CTTCTTGGTTTTTGGGGGGAATATGGATGCACAGACGGTGCAAACACAGAACTGGAAAAGC
AGATGGTCTGACTGTGCTATGGCCTCATCATCAAGACTTTCAATCCTATCCCAAGTGAAA
20 TAAATGGAATGAAATAATTCAAACACAAAAAAAAAAAAAAAAAAAAA

AB000520

GGATCCAAGCTATTGTCTGCCCATGGCTTCCCATCTCAGGACGCTCTCTGGCCGCTATC
ATCCCAGCAGTGAGTTTCAGCCCACTACTCTGAACCAGCCGCAGGTGGCTGCTATGGGAC
25 TGAAGCCATGAATGGTGCCGGCCCTGGCCCCGCGCAGCCGCCCGGTCCCAGTCCCGGT
CCCGGTCCCGGACTGGCGGCAGTTCTGCGAGCTGCATGCGCAGGCGGCCGCGCTGGACTT
TGCGCACAAGTTCTGCCGTTTCTGCGGGACAACCCAGCTTACGACACGCCCGACGCCGG
CGCTCCTTCTCCCGCCACTTCGCCGCCAACTTCCTGGACGTCTTCGGCGAGGAGGTGCG
CCGCGTGCTGGTGGCTGGGCCGACGACTCGGGGCGCGGCCGTGAGCGCAGAGGCCATGGA
30 GCCGGAGCTCGCGGACACCTCTGCACTCAAGGCGGCGTCTACGGCCACTCGCGGAGCTC
GGAGGACGTGTCCACGCACGCGGCCACCAAGGCCCGCGTTCGCAAGGGCTTCTCGCTGCG
CAACATGAGCCTGTGCGTGGTGGACGGCGTGCGCGACATGTGGCACCGGCGCGCCTCGCC
CGAGCCCGACGCGGCAGCTGCCCCGCGCACCGCCGAGCCCCGCGACAAGTGGACGCGGCG
CCTGAGGCTGTGCGGACGCTGGCTGCCAAGGTGGAGCTGGTGGACATTCAACGCGAGGG
35 GGCGCTGCGCTTCATGGTGGCCGACGACGCGGCCGCGGGCTCCGGGGGCTCGGCTCAGTG
GCAGAAGTGCCGCTGCTCCTGCGCAGGGCTGTGGCCGAGGAACGCTTCCGCCTGGAGTT
CTTCGTGCCGCCCAAAGCCTCCAGGCCCAAGGTGAGCATCCCACTGTGAGCCATCATTGA
GGTCCGCACCACCATGCCCCTGAAATGCCAGAGAAGGATAACACATTTCGTCTCAAGGT
AGAGAATGGAGCCGAATACATCTTGAGACCATCGACTCTCTGCAGAAGCACTCGTGGGT
40 AGCTGACATCCAGGGCTGCGTGGACCCCGGTGACAGTGAGGAAGACACCGAGCTCTCCTG
TACCCGAGGAGGCTGTCTGGCCAGCCGCGTGGCCTCCTGCAGCTGTGAGCTCCTGACTGA
TGCAGTCGACCTGCCCCGCCCCCAGAGACGACAGCCGTGGGTGCAGTGGTGACAGCCCC
CCACAGCCGAGGTGAGATGCCGTCAGAGAATCCCTGATCCACGTCCCGCTAGAGACCTT

TCTGCAGACCCTGGAATCCCCGGGCGGCAGCGGCAGTGACAGCAATAACACAGGGGAACA
GGGTGCAGAGACGGATCCCCGAGGCTGAACCCGAGCTGGAGCTATCCGACTACCCATGGTT
CCACGGGACACTGTCCCGGGTCAAGGCTGCTCAACTGGTTCTGGCAGGGGGGCCCCGGAA
CCACGGCCTCTTCGTGATCCGCCAAAGTGAGACTCGGCCTGGGGAGTACGTGCTGACCTT
5 CAACTTCCAGGGCAAGGCCAAGCACCTGCGCCTGTCCCTGAACGGCCACGGCCAGTGTC
CGTACAGCATCTGTGGTTCCAGTCTGTGCTTGACATGCTCCGCCACTTCCACACACACCC
CATCCCCTGGAGTCAGGGGGCTCGGCCGACATCACCTTCGCAGCTATGTGCGGGCCCA
GGACCCCCCACCAGAGCCGGGCCCCACGCCCCCTGCCGCGCCCGCGTCCCCGGCCTGCTG
GAGCGACTCGCCCGGCCAGCACTACTTCTCCAGCCTCGCCGCGGCCGCTGCCCGCCTGC
10 CTCGCCCTCCGACGCCCGCGGCCCTCCTCGTCTTCCGCCTCGTCGTCTCTGCGCGTC
GGGGCCCCGCCCCCGCGCCCCGTCGAGGGCCAGCTCAGCGCGCGGAGCCGCAGCAACAG
CGCCGAGCGCCTGCTGGAGGCCGTGGCCGCCACCGCCGCCGAGGAGCCCCCGGAGCCGC
GCCCCGCCGCGCGCGCGCCGTGGAGAACCAGTACTCCTTCTACTAGCCCCGCGCGCCGCC
CGGGTGGGACACGCCAAGCTCTTCAGTGAAGACACGATGTTATTAAAGCCTGTTTTAGG
15 GACTGCAAAA

AI820604 (IMAGE Clone Id: 1605108)

GATTCCAGCACGGGCTTCGCAGACTGCAGGACACAGAGGCACGCGTGACATCATGTCTT
CTAAGGAATTTGAACACTGTTGAGAAGACTGTGTACAAGAGAGATGTGCCATGTCAGCCT
20 TGCAAGGGACAGCGTGAAAACCTACCCATCTCCGGTCACCAAGTTGCAGGAGGCCAGGAGC
CAGGAGGGGAAACCGCTCAGTTTGCAAAACGTCGCTTCCACAAGCCTGATGGCTGAAACT
GCTCACTGTACCCTGAAACCAGCTTTACCTACAGCTTCTGAGATAAACTGCTGCAACTCT
GGGACCCACGATGCCTATCACAGTGGCTCATCAATGGAACCTGCCGGCTCCCAACCCTTC
CTAGGGCCCCTGAACTCTCTGAAAAGAGGAACAGAAATATTTCTCCTTTTTGTAAAATCT
25 TTAACCTTCCCTTTGTTCTTCATGTACACGCTGAACTGCAATTCTTCTTCCCAAATAAAA
CATTAAATTTAAAAAA

AI087057 (IMAGE Clone ID:1671188)

GGCCCCGGAGGGAGAGTAACCCGGCCCATCCATCCGTCGCCCCGGTTCTTGGGGAACTACT
30 TTCAGGGGCTTCTTGCCGTCCCCTCATCAGCTCTGTGCGAACCCTCTGTGCGCAGCCATT
GAGGAGACCCTGCCCCCTGGACCCTGACCACATATAGATTGAGGCCGAGGAGTGGCTGCC
CTGTCCCTTTTATGACAGCCCGCAGAAGCCCCGGGGTGAGGCATGGAGGAGGCAGGCGAC
AGCTGACAGGGACCCTGTTGGCCTCCAGCATGTCCAGCCAGCCGGGCAGGATTTCTCTGC
TTCTGGCTGGCAGCCAGGAACCTGAGTATGACAATGTTGTACTAAAGAAAGGCCCAAAGTG
35 ACAGAGGCAGCAGAGGGATGGTCCACCGCCCCCTTGCTTCTGCTGGTGACTCCTCCTGGC
CACTGCATCAGAAGAACCTCCTCTGCCCCCTTCTGGAGCCCGAGGCCTGGCCTGTCTTCGT
TGGGGCTGATAAATTGCCTCTCCCAGGGCCTGCTGGGTGAGTCACCATCCCAAAGCAGGA
AGGGTGCCCTGGAGAGAACCACCTCCTCCTACTCTTTTTTCCACTTCTCCTCTTTCTTT
CCCCAGCTGAGGAGGAACCTGGGGCATTTAGGGCAGAGGACAAAAGGATGTCAGCAATTG
40 CTTGGGCTGCTTGGCTATGCAAGCCTCCTGCCTGCTGATGGCCACTTCAGGGACAGCCTG
GGCCCAGGCACCCAGGGGGATGGCGGCAGCTTCTGCACCTTTCAGATTTCTTGGTGGCA
TTAAAGCATTTTCAGAACAAAAA

AJ272267

GGCGGGCCTGGACGGCCGCGTGCTGTACTGGCCACGCGGCCGCGTCTGGGGTGGCTCCTC
ATCCCTCAATGCCATGGTCTACGTCCGTGGGCACGCCGAGGACTACGAGCGCTGGCAGCG
5 CCAGGGCGCCCGCGGTGGGACTACGCGCACTGCCTGCCCTACTTCCGCAAGGCGCAGGG
CCACGAGCTGGGCGCCAGCCGGTACCGGGGCGCCGATGGCCCCGCTGCGGGTGTCCCGGGG
CAAGACCAACCACCCGCTGCACTGCGCATTCTTGAGAGGCCACGCAGCAGGCCGGCTACCC
GCTCACCGAGGACATGAATGGCTTCCAGCAGGAGGGCTTCGGCTGGATGGACATGACCAT
CCATGAAGGCAAACGGTGGAGCGCGGCCTGTGCCTACCTGCACCCAGCACTGAGCCGCAC
10 CAACCTCAAGGCCGAGGCCGAGAÇGCTTGTGAGCAGGGTGCTATTTGAGGGCACCCGTGC
AGTGGGCGTGGAGTATGTTAAGAATGGCCAGAGCCACAGGGCTTATGCCAGCAAGGAGGT
GATTCTGAGTGGAGGTGCCATCAACTCTCCACAGCTGCTCATGCTCTCTGGCATCGGGAA
TGCTGATGACCTCAAGAACTGGGCATCCCTGTGGTGTGCCACCTACCTGGGGTTGGCCA
GAACCTGCAAGACCACCTGGAGATCTACATTACAGCAGGCATGCACCCGCCCTATCACCCCT
15 CCATTACGACACAGAAGCCCCCTGCGGAAGGTCTGCATTGGTCTGGAGTGGCTCTGGAAATT
CACAGGGGAGGGAGCCACTGCCCATCTGGAAACAGGTGGGTTCATCCGCAGCCAGCCTGG
GGTCCCCCACCCGGACATCCAGTTCCATTTCCTGCCATCCCAAGTGATTGACCACGGGCG
GGTCCCCACCCAGCAGGAGGCTTACCAGGTACATGTGGGGCCCATGCGGGGCACGAGTGT
GGGCTGGCTCAAACCTGAGAAGTGCCAATCCCCAAGACCACCCTGTGATCCAGCCCACTA
20 CTTGTCAACAGAACTGATATTGAGGATTTCCGTCTGTGTGTGAAGCTCACCAGAGAAAT
TTTTGACAGGAAGCCCTGGCTCCGTTCCGAGGGAAAGAGCTCCAGCCAGGAAGCCACAT
TCAGTCAGATAAAGAGATAGATGCCTTTGTGCGGGCAAAGCCGACAGCGCCTACCACCC
CTCGTGACCTGTAAGATGGGCCAGCCCTCCGATCCCACTGCCGTGGTGGATCCGCAGAC
AAGGGTCCTCGGGGTGGAACCTCAGGGTCGTGCATGCCTCCATCATGCCTAGCATGGT
25 CAGCGGCAACCTGAACGCCCCCACAATCATGATCGCAGAGAAGGCAGCTGACATTATCAA
GGGGCAGCCTGCACTCTGGGACAAAGATGTCCCTGTCTACAAGCCCAGGACGCTGGCCAC
CCAGCGCTAAGACAGTTGCTGCTGGAGGATGACCAGGGAAGCCCCCTGATAAGCCAAGAG
GGCCAGCACAGCCCTTGCTCCCAGGCTCCTGCCTGAACTATCTAGCACACTAGGACCCA
GGTGGTACCCTACTCAGTGGCTGAGAATTGGATAAAGTCTTKGGGAAATGAGACAAGTAC
30 TGGGCAGTGAATCCAGCTCCTTTTCCCCAGCCTTTCCCTGTGGGCCATTTGGGGAAGGCC
AGCATTYCAGCCTGAGATGTTCTCCTGCCTCCTGGGGGGGCARAAGGGVTAGGWTGGT
TAACTCCTGCCGCATCCTTCCCTGCCTCCTGGAGGGACAGAAGGGGAGGATGGTTAACTC
CTGCCGCATCCTTTTTCTTGTGTTACGTGGCATTCTCTAACCAGGGCAGTGGTTCCTT
CCCAGGCCATGCACAGAGGCTGGGTGCCTGCCAGACCCACGGAGGGTTCGCGAAGGAAGG
35 GGCATCCTCCTTCTTGAGCTGCAAGCTTTAGCTGAGGCAGTAAGTCACACAGTAGTTAGT
TCAGCCTGGGCTGGCACATAAGTCCCCAGTGTCCCTGTTGAGAGGGGAAAGTTGCCTGCT
GGTTGAAAACTGGCTTTTCTCCTTCTCGCTGCCTAATTTCACTCTCAGAGTGAGGCAGGT
AACTGGGGCTCCACTGGGTCACTCTGAGAGGGTGTGGCTCTGGTCTTATTAAACCAGG
GCCAGGTGCAGGGCTCACACCTGTAATCCCAGCACTTTGGGAAGGTCACTTGAGCTCAGG
40 AGTTCAAGACCAGCCTGGGCAACATAGTGAGACCTTGTCTCTGGAAAACAATTAGCTGGG
CATGGTGGTACACACCTGTAGTCCCAGCTACTTGGGAGGCTGAGGCGGGAGGATGGCTTT
AGCCCAGGAGGTTGAGGCTCCTGTGAACCCTGATGGCACCACTGCACTCCAGCCTGGGTG
ACAGGGTGAGACCCTGTCTCAAAAAAAAAA

N30081 (IMAGE Clone ID: 258695)

CCGCCGTTGNCAAAGGGCCAGAAATATGGGCCATGGACNATCTCCATGCCTGGGGAAATT
CCCTCGGGTCTTTTGGNTAACNCCTTATAGAAAGGTAATGNCATGGAGTCTCTACAGGG
5 NGCACAAGGTGGACTAATTGATACGAAGAGCCCTGTAAATATGTGGGCAGCGGCAGATTT
TGACCATTTGGACCGAACTGTATTTGACACAGCGCAATATCTGGAAGTGGTTGGTCAAAA
ACCTGCTTGTCTTGTTAAATTTCTCTGTCCAAGGACATGGAATCTCTCTCTAATTTTAC
TTCAAATTTCCCTTTCCCTTCATTTCTCTAAAAACGTTAAATAAGAAAGAAGATTGTAAAG
CCAGCATTTGAAGCCTAAGTATTGAAAGTCTTTGACAATTTCTGAAATCAGACTTGACAT
10 CTTTCCCCCGCCTTGCAAATTTCTTGAAGAAATAAGAAGCTACATGTAAGCATCATCATG
TTTATTAAATTACAATGAGAACTCTCACTCAATCTTGACCAGAGCAGACTCTTAACCTTG
AAGCAGAGTCCCTCTAAAGGTAACCTTTGTGGTCACTCAATATTGTATTGGCATTTCAT
ATTAAATAGACATTTTCAGTAGCATTT

15 AI700363 (IMAGE Clone ID: 2327403)

TGGCCCGCGGTGCGGGTGGGATCCTAGCCCTGTCTCCTCTCCTGGGAAGGAGTGAGGGTG
GGACGTGACTTAGACACCTACAAATCTATTTACCAAAGAGGAGCCCGGGACTGAGGGAAA
AGGCCAAAGAGTGTGAGTGCATGCGGACTGGGGGTTGAGGGGAAGAGGACGAGGAGGAGG
AAGATGAGGTCGATTTCTGATTTAAAAAATCGTCCAAGCCCCGTGGTCCAGCTTAAGGT
20 CCTCGGTTACATGCGCCGCTCAGAGCAGGTCACCTTTCTGCCTTCCACGTCTCTCTTCAAG
GAAGCCCCATGTGGGTAGCTTTCAATATCGCAGGTTCTTACTCCTCTGCCTCTATAAGCT
CAAACCCACCAACGATCGGGCAAGTAAACCCCTCCCTCGCCGACTTCGGAAGTGGCGAG
AGTTCAGCGCAGATGGGCCTGTGGGGAGGGGGCAAGATAGATGAGGGGGAGCGGCATGGT
GCGGGGTGACCCCTTGGAGAGAGGAAAAAGGCCACAAGAGGGGCTGCCACCGCCACTAAC
25 GGAGATGGCCCTGGTAGAGACCTTTGGGGGTCTGGAACCTCTGGACTCCCCATGCTCTAA
CTCCCACACTCTGCTATCAGAACTTAACTTGAGGATTTTCTCTGTTTTTCACTCGCAA
TAAATTCAGAGCAAACAAAAAAAAAAAAAAAAA

AL117406

30 CAATAGGCCGGCTTTTGAACTGCTTCGCAGGGGACTTGGAACAGCTGGACCAGCTCTTGC
CCATCTTTTCAGAGCAGTTCCTGGTCCTGTCCTTAATGGTGATCGCCGTCTGTGATTG
TCAGTGTGCTGTCTCCATATATCTGTTAATGGGAGCCATAATCATGGTTATTTGCTTCA
TTTATTATATGATGTTCAAGAAGGCCATCGGTGTGTTCAAGAGACTGGAGAACTATAGCC
GGTCTCCTTTATTCTCCACATCCTCAATTCTCTGCAAGGCCTGAGCTCCATCCATGTCT
35 ATGGAAAACTGAAGACTTCATCAGCCAGTTTAAGAGGCTGACTGATGCGCAGAATAACT
ACCTGCTGTTGTTTCTATCTTCCACACGATGGATGGCATTGAGGCTGGAGATCATGACCA
ACCTTGTGACCTTGGCTGTTGCCCTGTTTCGTGGCTTTTGGCATTTCCTCCACCCCCTACT
CCTTTAAAGTCATGGCTGTCAACATCGTGCTGCAGCTGGCGTCCAGCTTCCAGGCCACTG
CCCGGATTGGCTTGGAGACAGAGGCACAGTTCACGGCTGTAGAGAGGATACTGCAGTACA
40 TGAAGATGTGTGTCTCGGAAGCTCCTTTACACATGGAAGGCACAAGTTGTCCCCAGGGGT
GGCCACAGCATGGGGAAATCATATTTAGGATTATCACATGAAATACAGAGACAACACAC

CCACCGTGCTTCACGGCATCAACCTGACCATCCGCGGCCACGAAGTGGTGGGCATCGTGG
GAAGGACGGGCTCTGTAGGTTTTTACTGAGCACCTACTATGTGCCTGGGAACCGAAAGGG
AAGTCCTCCTTGGGCATGGCTCTCTTCCGCCTGGTGGAGCCCATGGCAGGCCGGATTCTC
ATTGACGGCGTGACATTTGCAGCATCGGCCTGGAGGACTTGCGGTCCAAGCTCTCAGTG
5 ATCCCTCAAGATCCAGTGCTGCTCTCAGGAACCATCAGATTCAACCTAGATCCCTTTGAC
CGTCACACTGACCAGCAGATCTGGGATGCCTTGGAGAGGACATTCTTGACCAAGGCCATC
TCAAAGTTCCCCAAAAAGCTGCATACAGATGTGGTGGAAAACGGTGGAACTTCTCTGTG
GGGGAGAGGCAGCTGCTCTGCATTGCCAGGGCTGTGCTTCGCAACTCCAAGATCATCCTT
ATCGATGAAGCCACAGCCTCCATTGACATGGAGACAGACACCCTGATCCAGCGCACAATC
10 CGTGAAGCCTTCCAGGGCTGCACCGTGCTCGTCATTGCCACCGTGTCAACACTGTGCTG
AACTGTGACCACATCCTGGTTATGGGCAATGGGAAGGTGGTAGAATTTGATCGGCCGGAG
GTACTGCGGAAGAAGCCTGGGTCAATTGTTTCGAGCCCTCATGGCCACAGCCACTTCTTCA
CTGAGATAAGGAGATGTGGAGACTTCATGGAGGCTGGCAGCTGAGCTCAGAGGTTACAC
AGGTGCAGCTTCGAGGGCCACAGTCTGCGACCTTCTTGTTTGGAGATGAGAACTTCTCCT
15 GGAAGCAGGGGTAAATGTAGGGGGGGTGGGGATTGCTGGATGGAAACCCTGGAATAGGCT
ACTTGATGGCTCTCAAGACCTTAGAACCCAGAACCATCTAAGACATGGGATTGAGTGAT
CATGTGGTTCTCCTTTTAACTTACATGCTGAATAATTTTATAATAAGGTAAAAGCTTATA
GTTTTCTGATCTGTGTTAGAAGTGTGCAAATGCTGTACTGACTTTGTAAAATATAAAAC
TAAGGAAACTCAAAAAAAAAAAAA

20

M92432

CCCACAGGGGGACCGGCCCTGTGACCCCTCACCGGGGCGCTGGGCCCCGAGCCCCGGACTT
CCCTAAGCCGGCAATGACCGCCTGCGCCCGCCGAGCGGGTGGGCTTCCGGACCCCGGGCT
CTGCGGTCCCGCGTGGTGGGCTCCGTCCCTGCCCGCCTCCCCCGGGCCCTGCCCGGCT
25 CCCGCTCCTGCTGCTCCTGCTTCTGCTGCAGCCCCCGCCCTCTCCGCCGTGTTACGGT
GGGGTCTTGGGCCCCCTGGGCTTGCAGCCCCATCTTCTCTCGGGCTCGCCCGGACCTGGC
CGCCCGCCTGGCCGCCCGCCGCTGAACCGCGACCCCGGCCTGGCAGGCGGTCCCCGCTT
CGAGGTAGCGCTGCTGCCCCGAGCCTTGCCCGACGCCGGGCTCGCTGGGGGCGGTGTCTCTC
CGCGCTGGCCCCGCGTGTGCGGCCTCGTGGGTCCGGTGAACCCTGCGGCCTGCCGGCCAGC
30 CGAGCTGCTCGCCGAAGAAGCCGGGATCGCGCTGGTGGCCTGGGGCTGCCCTTGACGCA
GGCGGAGGGCACCACGGCCCCCTGCCGTGACCCCGCGCGGATGCCCTCTACGCCCTGCT
TCGCGCATTCGGCTGGGCGCGCGTGGCCCTGGTCACCGCCCCCAGGACCTGTGGGTGGA
GGCGGGACGCTCACTGTCCACGGCACTCAGGGCCCCGGGGCTGCCTGTGCGCTCCGTGAC
TTCCATGGAGCCCTTGGACCTGTCTGGAGCCCGGGAGGCCCTGAGGAAGGTTCCGGACGG
35 GCCCAGGGTCACAGCAGTGATCATGGTGATGCACTCGGTGCTGCTGGGTGGCGAGGAGCA
GCGCTACCTCCTGGAGGCCGAGAGGAGCTGGGCCTGACCGATGGCTCCCTGGTCTTCTCT
GCCCTTCGACACGATCCACTACGCCTTGTCCCCAGGCCCGGAGGCCTTGGCCGCACTCGC
CAACAGCTCCCAGCTTCGAGGGCCACGATGCCGTGCTCACCCTCACGCGCCACTGTCC
CTCTGAAGGCAGCGTGCTGGACAGCCTGCGCAGGGCTCAAGAGCGCCGCGAGCTGCCCTC
40 TGACCTCAATCTGCAGCAGGTCTCCCCACTCTTTGGCACCATCTATGACGCGGTCTTCTT
GCTGGCAAGGGGCGTGGCAGAAGCGCGGGCTGCCGAGGTGGCAGATGGGTGTCCGGAGC
AGCTGTGGCCCGCCACATCCGGGATGCGCAGGTCCCTGGCTTCTGCGGGACCTAGGAGG
AGACGAGGAGCCCCCATTCGTGCTGCTAGACACGGACGCGGCGGGAGACCGGCTTTTTTG

CACATACATGCTGGATCCTGCCCCGGGGCTCCTTCCTCTCCGCCGGTACCCGGATGCACCTT
CCCCGCTGGGGGATCAGCACCCGGACCTGACCCCTCGTGCTGGTTCGATCCAAACAACAT
CTGCGGTGGAGGACTGGAGCCGGGCCTCGTCTTTCTTGGCTTCCTCCTGGTGGTTGGGAT
GGGGCTGGCTGGGGCCTTCCTGGCCCATTATGTGAGGCACCGGCTACTTCACATGCAAAT
5 GGTCTCCGGCCCCAACAAGATCATCCTGACCGTGGACGACATCACCTTTCTCCACCCACA
TGGGGGCACCTCTCGAAAGGTGGCCCAGGGGAGTCGATCAAGTCTGGGTGCCCCGAGCAT
GTCAGACATTCGCAGCGGCCCCAGCCAACACTTGGACAGCCCCAACATTGGTGTCTATGA
GGGAGACAGGGTTTGGCTGAAGAAATTCCCAGGGGATCAGCACATAGCTATCCGCCCAGC
AACCAAGACGGCCTTCTCCAAGCTCCAGGAGCTCCGGCATGAGAACGTGGCCCTCTACCT
10 GGGGCTTTTCTGGCTCGGGGAGCAGAAGGCCCTGCGGCCCTCTGGGAGGGCAACCTGGC
TGTGGTCTCAGAGCACTGCACGCGGGGCTCTCTTCAGGACCTCCTCGCTCAGAGAGAAAT
AAAGCTGGACTGGATGTTCAAGTCCTCCCTCCTGCTGGACCTTATCAAGGGAATAAGGTA
TCTGCACCATCGAGGCGTGGCTCATGGGCGGCTGAAGTCACGGAAGTGCATAGTGGATGG
CAGATTTCGTACTCAAGATCACTGACCACGGCCACGGGAGACTGCTGGAAGCACAGAAGGT
15 GCTACCGGAGCCTCCCAGAGCGGAGGACCAGCTGTGGACAGCCCCGGAGCTGCTTAGGGA
CCCAGCCCTGGAGCGCCGGGGAACGCTGGCCGGCGACGTCTTTAGCTTGGCCATCATCAT
GCAAGAAGTAGTGTGCCGCAGTGCCCTTATGCCATGCTGGAGCTCACTCCCGAGGAAGT
GGTGCAGAGGGTGCAGGAGCCCCCTCCACTGTGTGCGGCCCTTGGTGTCCATGGACCAGGC
ACCTGTTCGAGTGTATCCTCCTGATGAAGCAGTGTGGGCAGAGCAGCCGGAAGTTCGGCC
20 CTCCATGGACCACACCTTCGACCTGTTCAAGAACATCAACAAGGGCCGGAAGACGAACAT
CATTGACTCGATGCTTCGGATGCTGGAGCAGTACTCTAGTAACCTGGAGGATCTGATCCG
GGAGCGCACGGAGGAGCTGGAGCTGGAAAAGCAGAAGACAGACCGGCTGCTTACACAGAT
GCTGCCTCCGTCTGTGGCTGAGGCCTTGAAGACGGGGACACCAGTGGAGCCCCGAGTACTT
TGAGCAAGTGACACTGTACTTTAGTGACATTGTGGGCTTACCACCATCTCTGCCATGAG
25 TGAGCCCATTGAGGTTGTGGACCTGCTCAACGATCTCTACACACTCTTTGATGCCATCAT
TGGTTCCACGATGTCTACAAGGTGGAGACAATAGGGGACGCCTATATGGTGGCCTCGGG
GCTGCCCCAGCGGAATGGGCAGCGACACGCGGCAGAGATCGCCAACATGTCACTGGACAT
CCTCAGTGCCGTGGGCACTTTCCGCATGCGCCATATGCCTGAGGTTCCCGTGCGCATCCG
CATAGGCCTGCACTCGGGTCCATGCGTGGCAGGCGTGGTGGGCCTCACCATGCCGCGGTA
30 CTGCCTGTTTGGGGACACGGTCAACACCGCCTCGCGCATGGAGTCCACCGGGCTGCCTTA
CCGCATCCACGTGAACCTTGAGCACTGTGGGGATTCTCCGTGCTCTGGACTCGGGCTACCA
GGTGGAGCTGCGAGGCCGCACGGAGCTGAAGGGCAAGGGCGCCGAGGACACTTTCTGGCT
AGTGGGCAGACGCGGCTTCAACAAGCCCATCCCCAAACCGCCTGACCTGCAACCGGGGTC
CAGCAACCACGGCATCAGCCTGCAGGAGATCCACCCGAGCGGCGACGGAAGCTGGAGAA
35 GGCGCGGCCGGGCCAGTTCTCTTGAGAAGTGAGGCCCGGCCCGGACAGGGTCTGGGCCC
TGCTCCCTGTCCCATCTGCAGTGGACCCAGGCACCCCCCTTTGAGGAGGTGGGGTGAAC
TGCTCCTTGGCAGGGATTTGTGACACTGCATTGCTGGGCTGTGTTCTCAGGGCTCTTCTG
GACCTTGCACCGTGGATAACCAGGCCATGTGCCATGGTATTTGGGTCCTGGGAGGGTGGGT
GAAATAAAGGCATACTGTCTT
40

AL050227

CTTTCACAGAAAGAAAGTAACAGGCATAATTCCTGTTGATGAGGCTGGGATTGTTTTTAA
GAGGAGAGATAATAACTTCATATTTTTAAAGTGCCAGTAGCCTAATATGTGAAACAGATC

AGAATCTGTTGTGTAGTAAGTCTGCTTTGTTGAAGAATTTATTATGGGAGTAAAGATAAG
AAGGAAAGAGATCACCATCAGAAACAAGTCAGCCTTTTCATGCTTTTTTGAGCATTTTTG
GAGATGATTCCACTTCTCAAGTTATTATCATTTGTGCATCTCTTCAATGCTATTGTTAAA
TGCTTTAGAATTAGAATATTTTGATCCTTTAATTAAAGTAAGCCAAACGTCTAGGCAAAA
5 ACAGCCAATCATTAAACTTTAATAGTAATTCAAATATAGATTTCTCATACAGTTTTCCAT
GTCTGTAGAAATCAAAGTTGTAATGTTAAGCAGAGGGAAATGCGTGTGATTTACTAATAC
ACTTCAACGTTCTACTTTTGAAAGGATACTCATGTGGGTGGGGCAGAGAACATAGAAAAA
GATATGATGGAAAACCTGTCCATTTTCTACCTGTTAACCTTCATCATTTTGTGCAGGCCC
TGGAAGCAAAGAGAGGAAGGGACCGACTGCATTTATCTTTGAACACTTGAGCATCAGTAG
10 TACTACTGAGTGGCCAGGGTCTTGTCTGTCAAAGCAAATGATAAGTTCACTCAGGCCAT
TATTGACTGCTGAACTCTCTTCCCTCCCAACTCTTCTTGAAAGAGAAAAAATACTTTG
CCTTCTTGCTCTCCTTATCAAATGTTTTTGTACAAATAGTGTAAAGCCTGTTTAAGCAAAC
CAATTAATAATAGGCACTGATTATTTTGATCTGTTTGTAAACAAATGAATGTAAGTACTATT
TACATGGTGTGCCTAGGAGGAGCTGAAATCATTGGCACTTTAATCCATATTGTAAAGATC
15 AGTATCAAAAGCATAGTGTCTTCACCTCTCCTCCTCAGCATCCATCTCTATATACTTGA
TTAAATGGAAAAGTCTCTTTTATCACCTCTATGTAAAGTTTTATGGGTAGTTATCGTCAG
TGTATTTAAATATATCTTCTAGTATGTTTTAAAGGCTGGTCTTCAATACTGTGGAGACAA
AAAATAAAAGAGCGTATGAAAAGTACGTTAGACTTTTGCTGGCATTCAAGTCATGGCTAG
TCTGTGTATTTAATAAATGTGTGTTATTTATGTCGTGTTTGTCAATGGAAAATAAAGTTG
20 AATATTCTGAAAAAAAAAAAAAA

AW613732 (IMAGE Clone ID: 2953502)

CCTANAAGTNCCATTTTGGCAAGGATAAACTCCCATGACAANCTCCCANTACTGCATGTG
AATGAATAAGAAACAAGAANTGACCACACCAAAGCCTCCCTGGCTGGTGTTACANGGGAT
25 CAGGTCCACAGTGGTGCAGATTCAACCACCACCCAGGGAGTGCTTGCAGACTCTGCATAG
ATGTTGCTGCATGCGTCCCATGTGCCTGTCAGAAATGGCAGTGTTAATTCTCTTGAAAGA
AAGTTATTTGCTCACTATCCCCAGCCTCAAGGAGCCAAGGAAGAGTCATTACATGGAAG
GTCCGGGACTGGTCAGCCACTCTGACTTTTCTACCACATTAAATTCTCCATTACATCTCA
CTATTGGTAATGGCTTAAGTGTAAGAGCCATGATGTGTATATTAAGCTATGTGCCACAT
30 ATTTATTTTGTAGACTCTCCACAGCATTATGTCAATATGGGATTAATGCCTAACTTTGT
AAATATTGTACAGTTTGTAATCAATGAATAAAGGTTTTGAGTGTAATAAAAAAAAAAAAAA
AAAAAA

BC007783 (IMAGE Clone ID: 4308472)

GGCACGAGGGCAAAGAGTAGTCAGTCCCTTCTTGGCTCTGCTGACACTCGAGCCCACATT
CCATCACCTGCTCCCAATCATGCAGGTCTCCACTGCTGCCCTTGCCGTCCTCCTCTGCAC
CATGGCTCTCTGCAACCAGGTCCTCTCTGCACCACTTGCTGCTGACACGCCGACCGCCTG
CTGCTTCAGCTACACCTCCCGGCAGATTCCACAGAATTTTCATAGCTGACTACTTTGAGAC
GAGCAGCCAGTGCTCCAAGCCCAGTGTCATCTTCCTAACCAAGAGAGGCCGGCAGGTCTG
40 TGCTGACCCAGTGAGGAGTGGGTCCAGAAATACGTCAGTGACCTGGAGCCGAGTGCCTG
AGGGGTCCAGAAGCTTCGAGGCCAGCGACCTCAGTGGGCCCAGTGGGGAGGAGCAGGAG
CCTGAGCCTTGGGAACATGCGTGTGACCTCCACAGCTACCTCTTCTATGGACTGGTTATT

GCCAAACAGCCACACTGTGGGACTCTTCTTAACTTAAATTTTAATTTATTTATACTATTT
AGTTTTTATAATTTATTTTTGATTTCACAGTGTGTTTGTGATTGTTTGCTCTGAGAGTTC
CCCCGTGTCCCCTCCACCTTCCCTCACAGTGTGTCTGGTGACAACCGAGTGGCTGTCATCG
GCCTGTGTAGGCAGTCATGGCACCAAAGCCACCAGACTGACAAATGTGTATCAGATGCTT
5 TTGTTTCAGGGCTGTGATCGGCCTGGGGAAATAATAAAGATGTTCTTTTAAACGGTAAAA
AAAA

X81896

AGAAAACATATTTTCTAAATATTAACACTGAAAATGTTTTGTTAGCTTTTCCTTCTTTCTC
10 TCCAGAAGAAACATGGATAGATGATAGCTGTTTCATTGTTTGTGTTTGTCAAGCATATTC
ACTTTCCTCCTTGTCTCTGATTCTGAGCAAAGGGCCTCAGACTCTGAACTTCCCTCAAG
TGCCGTTGTTATGTGAACTCTTCCATTCCAGAGAGGTTCTCATGCTCCCCCCCC
CTCCTTATTTGTAGCAATCGTAGCAACTAATTCCTAAGTACAAGGGAGTTTTTTACAC
TCCTCCATTTTTATAGCATCTGCATTTTTTTTTTTTTGTTAGGTACATGTATACACCTGCC
15 TGAGTATAAATACTCTCTCTACCTAATAATAACATCAACCAACATCTTTTCCAAATTAGG
GCCACAGAACAGCAACATTTGTCTGACAGTAGTATAAAGAATAATGATAGCTCTATCCTT
AAGAAGTATTTCTTTCTTTTATATAGTCCCGTTAGGGTTTAAAACCATATTGATCAA
CTAGAAAGAAAAATATGAAAAGAGAAAAATATTTTAATTTAAAAATTGTAATACATTGAT
TTATAAAATGCCTTCTCTGATACTTTTGAAACAGATGTGAAAACAGAAAAAGAAAAAAT
20 TGTCTGAAATGTTTATTTTGCAAAACAGTGCAATAGAATCTAGTTATGCCTTCATCACTG
TTGACAGTAAATACTGACAGCCCCCTTGCAGTGTGTTAGTTTTAGATCACTCTGTTTTAGT
TGAGAGAAATGTTTTATATCATGGTTTTTATATGAATACAAATTATTTCTCAAAGATTTA
TAGCACACACTATTCTCAGGAATTCTGTATTACATGAATGCTGCTTATATATTTTCATAT
TCTAACTTGTCTTTTCAAGCAAATACTAATATATATGTGCATGCAGTCTGCCTTGACAA
25 GTTGTTCGAAGCTGAAGAGCTTTCAGTGTACAATGTGTGGAAAAATCACCATAGATCATGG
CTGAAATAGTTTGTAATTGTCTGAGTCTGTGCACGTAATTTTAGATAAAATGCTGCTGAG
TGACTGCATGATGAGATACAACTTCTGAATGCTGCACATTCTTCCAAAATGATCCTTAGC
ACAATCTATTGTATGATGGAATGAATAGAAAACCTTTTTCACTCAATAAATTATTATTTGA
TATGGTAAAAA
30

BC004960 (IMAGE Clone ID: 3632495)

CCCAAGGTTGTTATATCTTCATGTCCTCATTCTTAGGGAGGTACCTTCAGAACCAATAG
TGACCCCTAACTTCTCTGGTGGTTCGTTCCATGAAAGGCAAAGGAGTGTGAGAGAGGAGT
GGATGGTCAACCTCCCACTGCCATGGTAACATGGGTGCTGGCTGATGGGAGCAGAAAATA
35 ATTTAGTGAAAGTCTGTGGGGGCAGTCACAAGATGTCTGAGAAAACCTGGCGAGCCAGCTG
CTGAAAACAGGGACAAGGAAGCCTCCGTGGCTGGAGCCCAAATCACACTGCAGACCCAGA
CACCGTGACCACCACCATGGACTCCAGAGAGAGCAGCTTATAGTACTCAATCAGCTGCCA
CTACCACCATCCAGAACACCAGATGTTGTAGCCATGGCTGCAGCAGGAATGGATGTCCCA
CTGTCCCTGCTCCTCGGTGTGACTTGCTCCCAAGTTCAGGGCAGGTCCATCTGATTGGCT
40 GAGTCTGGAATGTCTGCCTGTGCCTCAGCTGTGAGGGAGGCAGGGAAAGTAAGCCTTTTC
AGCTTCTGTCTGGGAGGTGGGCTCTGCCTCCTACCAAGAATCAAAGGTGGAGGATCTT
CAAACACAGGAAAAAGAACCCGGATCCTGGCACCCCAAATTTTCAGAGTCCATTTTCAGAG

CATAAGAAATTGAGGGTCCAAGATCATTCATGTAAGAAGTTTAGAGGGGGAAGAAAAGAA
TGATAAACGAAAAGAACAGCAATAGTAAAGGATCTTTTCTTTGTTTCAGTAAGATGAAGA
GGCCTGAGCAGTTTCGTGGAGGGGAAGAAACAGGAAAACCTCTTCAAAGACAAAAGCT
GGCACTGCATTCTCTCTCTGTAGCAGGACAGAACTGTCTAAAGACAAGACCCCTTTGGCC
5 AAAATAAAGGAACCTGAAACATTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAACCTCGGG

AK027250

AAATTTAATTAATTATAAACTCAGTCTCTTGGTTGCACCAGCCACATTTTCAGATGCTCAA
10 TAGCCACATGTGGCTAGTGGGTACCATATTGGACAGGGCAGCTATAGAATATTTCCATCA
TTGCAGAAAGTTCTATTGGATAGTACCATAATCTTTTTATAGTAACTTGGAATACTATT
TGATATTAGATGTTAGACCACAAAAAGAAGAAAAATGTTAGGACTATTTTCAGATATAAAA
AGGAACTGAATTGTGACATAATTAGCATCTTACATTCCATACAGTTGAATACCTTATGCT
GTGACAACCATAGTTAATCATTTTCAGTGCTGTTCAACATACATACCTATCAGCAGTGTGT
15 TTAGACCAGGGTCTGCAAACCTTCTGTGAATGGACAAAGAGTAAATACTTTAGTAAATG
TCTTAGGCTTTGTGGCCTACATGATCTTTGTTGCAAGTACTCAACTCTGCCATTATAGAG
TTAAAGCAGCCATACACAATATATAAACAAAATGGGCATAGTTGTATTTTCAGTAAACTT
TATTTACAAAGACAGGCGGTAGGCCAGATTTGGCTTGCATGCTGTAGAGCTGTGGTCTAA
ATTTTATTCATAGACTTTCTTTGCAAATACAGTGTGAGTATTGTTCCATTTACAGTATTA
20 TTATTTTTTTAGATACCTGGTTTTTAGATTCTTGCCTGGTAACTTTTTACTGAAAATACAA
GAATTTTCGTACTGCATTTGCATCTCCGAGATTAGGGAGCACCTGTCAGGATATGTTGTTT
TATCAGGGTTACTTCTGTTGACTACCTCTTAGATTTTGATACAGTTATATTGTTGAGTTT
CATTTTCATATATTCTTGTAGTGTCTGCTTGCCTGTGACTTCTGGTAAAATAAAATAAGC
CTTTGAAAATATTTTAGCATGGTATTTAACATTTTCTAAATATTATGGCATTTTGACATA
25 TTTTAGTCAGCGAAGACATCTGCCCTTTGGTGTCTTCTACTTGCTTATGATTGAGATTTT
ACAAGCCCTTCAAACCTCCGTTTTAAAGGAATTTATTGTAAAACATTAACTTTAATAAATT
AGTGTTTTTACAGATCAGATCATTATACTTGGAACCTTCTAAATCATGCAATTTCTGAATA
AGGACATAAGGCTAGATTCAATTTTCTTAATAGAGAAAAAGGAAATTTCTGATTTATCAC
TTTTCTAGTTGATAAGTAGGATTCAAACGTTTGATATGTAAGTATTTATATAAGACTAA
30 TGTAATTTAAAGTTCTGTATTATTGTGATTAATCATACAGAAATTCAGGAACTGATCAGA
AGTGAGATTCTTTTCCACATCTGGTTAATGTAGTGAGTTGACACCCTGTGGGTGGTAAAG
CATTATAAACATTTTCATCTTGAACCATGATTTATACACATCTGTGTTATAAGGGAGGCTT
GAGTACATATACCAATGAAGAGATATTCAGCATTTGTCTATTTGATAAGGAATTAAATGT
CCTAGTGATTATAAAGTAAACACAGACCAATTTGCAAATGATCTTCAATGTTAAGCAC
35 TTGCTCTAAGATTAAAATTCCTTTTCTTTTAAAGGTAAAGGTGTGTACGTATGGCAGTG
ATGTCTATGTTGAGATTAACTTATGTATTGAGGAAAATTTGAAGTTTATTTTTTCGATGA
ATAAGGCTGTCAAATGATTTAGTATAGATTAATGACATCTTTTTTAGAAATATTAAAGTG
AGTATTCCTCATTATGTCATCATTTCTGATAATTAGAGTGCTAATTTGAATGTTAGATAA
TGTTTCCACATCTATACCTATTTCTTTCTAGGGCACCTCTGACCCTGGGGCTTGGGGATG
40 GCCTTTAGGCCACAGTAGTGTCTGTGTTAAGTTCACTAAATGTGTATTTAATGAGAAACA
TTCCTATGTAAAATGTGTGTATGTGAACGTATGCATACATTTTTATTGTGCACCTGTAC
ATTGTGAAGAAGTAGTTTGGAAATTTGTAAAGCACAAACCATAAAAGAGTGTGGAGTTAT
TAAATGATGTAGCACAAATGTAATGTTTAGCTTATAAAAGGTCCTTTCTATTTTCTATGG

CAAAGACTTTGACACTTGAAAAATAAAACCAATATTTGATTTATTTTTGTAAGTATTTAG
GATATTATTTTAAATAAATGATTGTCCATTATCAATAAAAAAAAAAAAAAAAAAAAA

Sequences from Table 5 not disclosed above

5

NM_014298

GTCCTGAGCAGCCAACACACCAGCCCAGACAGCTGCAAGTCACCATGGACGCTGAAGGCC
TGGCGCTGCTGCTGCCGCCCGTCACCCTGGCAGCCCTGGTGGACAGCTGGCTCCGAGAGG
ACTGCCCAGGGCTCAACTACGCAGCCTTGGTCAGCGGGGCAGGCCCTCGCAGGCGGCGC
10 TGTGGGCCAAATCCCCTGGGGTACTGGCAGGGCAGCCTTTCTTCGATGCCATATTTACCC
AACTCAACTGCCAAGTCTCCTGGTTCCTCCCCGAGGGATCGAAGCTGGTGCCGGTGGCCA
GAGTGGCCGAGGTCCGGGGCCCTGCCCACTGCCTGCTGCTGGGGGAACGGGTGGCCCTCA
ACACGCTGGCCCCGCTGCAGTGGCATTGCCAGTGTGCTGCCGCCGCTGCAGTGGAGGCCGCCA
GGGGGGCCGGCTGGACTGGGCACGTGGCAGGCACGAGGAAGACCACGCCAGGCTTCCGGC
15 TGGTGGAGAAGTATGGGCTCCTGGTGGGCGGGCCGCCTCGCACCGCTACGACCTGGGAG
GGCTGGTGTATGTTGAAGGATAACCATGTGGTGGCCCCCGGTGGCGTGGAGAAGGCGGTGC
GGGCGGCCAGACAGGCGGCTGACTTCGCTCTGAAGGTGGAAGTGGAATGCAGCAGCCTGC
AGGAGGTCGTCCAGGCAGCTGAGGCTGGCGCCGACCTTGTCCTGCTGGACAACTTCAAGC
CAGAGGAGCTGCACCCACGGCCACCGCGCTGAAGGCCCAGTTCCCGAGTGTGGCTGTGG
20 AAGCCAGTGGGGGCATCACCTGGACAACCTCCCCAGTTCTGCGGGCCGCACATAGACG
TCATCTCCATGGGGATGCTGACCCAGGCGGTCCCAGCCCTTGATTTCTCCCTCAAGCTGT
TTGCCAAAGAGGTGGCTCCAGTGCCCAAAATCCACTAGTCCTAAACCGGAAGAGGATGAC
ACCGGCCATGGGTAAACGTGGCTCCTCAGGACCCTCTGGGTACACATCTTTAGGGTCAG
TGAACAATGGGGCACATTTGGCACTAGCTTGAGCCCAACTCTGGCTCTGCCACCTGCTGC
25 TCCTGTGACCTGTGAGGGCTGACTTCACCTCTGCTCATCTCAGTTTCCTAATCTGTAAAA
TGGGTCTAATAAAGGATCAACCAAAAAAAAAAAAAAAAAAAAA

AF033199

CGGGGCATGCTGCTTCCCTTCACCTTCCACCATGATTGTAAGTTTCCTGAGGCCTCCCCA
30 GGTGTGCTTCTGTACAGCCTGTGGAATGTTACCAAAGACGTTGGAAGAGGTGGCTATGGG
ACATCACCTGGGAGAAAGTGAAGCAAATGGACACTGTTTCAAGTCCATATACAGAAACA
TACTTGGA AAAAATATAGAAACCTGGTTTTTGCTAGATGGGAAGCTTGCAGCTGGGGCCAAG
ACATCAAGAGTAGAGCAGCAGGACATTTCAAAGAAGATTAACTCAAAGATTAGAGATGG
AAGAACTTGCAAAGAGAAAGTCTGTACCGGAAGAAATCTGGAAATCTAGAGGCCAGTTTA
35 AGAATCAGCAGCTAAACAAGGAGAATAATCTAGGGCAAGAGATAGCTACCTGCACAAAAA
TTCCTACCAGAAAAAGAGACATAGAATCTAATGAATTTGTGAAAAATTTTACTGTAAGAT
CAATACTTGTTGCAGAACAGATAGATCCTATGGAAGAGAATTGTCATAAATATGGTACAT
GTTGAAAGATGCTCAAACAAAACCTCAGATTTAATTATACAAAGAAAGTATGATGGAAAAA
AAAAAACCTTGTAATATAGTGAATGTGGGAGAACCTTCAGAGGCCACATCACTCTTGTT
40 CAGCATCAAATAACTCATTGTGGAGAGAGACCCTGTAAATGTACTGAGTGTAGAAAGGGA
TTAATCAGAGTTCCCACTTAAGAAATAATCAGAGAAAAACTCTTTCAGGAGAAAAGCCC

TACAAATGCAGTGAGTGTGGGAAGGCCTTCAGTTATTGCTTAGTTCTTAATCAACACCAG
AGAATTCACAGTGGAGAGAAACCTTATGAGGGTACTGAATGTGGCAAGACATTCATTCAG
TCGTACATACCTTACTCAGCATCAAAGAATTCACACACTGGTGAGAAGCCCTATACATGT
CTTGAATGTGGAAGGCTTTTTAGTCAGAACACACATCTTACTCTACATCAGAGAATCCAT
5 ACTGGAGAGAAACCTTATGAATGCAATGAATGTGGTAGGTCCTTTAGTCAGACTGCACAT
CTTACTCAACATCAAAGAATGTATACAGGAGAAAACTCTATGAATGTAATGAATGTGAG
AAAGCCTTCCATGATCACTCAGCTCTTATTCAACATCATATTGTCCATACTGCAGAGAAA
CCCTATGATATCATGACTGGGAAAACCTTTCAGTTACTGTTTCAGACCTCATTCAACATCAG
AGAATGCACACTGGAGAGAAACCATACAAATGCAATGAATGTGGGAATGCCTTTTAGTGAT
10 TGTTTCATCCCTTATTTCAGCATCAAAGAACTCACACTGGAGAAGAGCCTTATGAATGTAAG
CAATGTGGAAAAGCCTTTAGCAGAAGCACATACCTTACTCAACATCAGAGAAGTCACGCA
GGAGAGAAACAGTATAAATGCAATGAATGTGAGAAAACCTTTCAGCCTGAGTTCATTTCCTT
ACACAGCATATGAGGGTTCAGACTGGAGAAAAACCTTACAAATATAATGAATATGGAAAA
GCTTTTAGTGACTGCTCAGGACATTTTCAGAGAACTCACACTGGAGAGAAGCCCTGTGAA
15 TGTAATGACTGTGGGAAACCTTTCAGTTTCTGTTTCAGCCCTAATTCAACATAAGAGAATT
CATAACCAGAAAGAAGCCCTGACTGTACCTTCATACCAGTAAATGCACTGACTGTGGAAAA
GCCTTCAGTGATTGGTTAGCACTTGTTCAACATCAGATAACTCAACACTGGAGAAAAACC
GTATAAATGTACTGAATGTGGAAAAGCCTTCAGTTGGAGTACAGACCTCAAAAATCACCA
GAAAACCTCATACTAGTGAAAAATCCTATAAATGTAATGAATGTAGAAAGGCCTTTAGTTA
20 CTGCTCTGGTCTTATTCAATGTCAGGTCATTCATACTATAGAAAAACCTTATGAATACGG
TAAATGTGGCAAAGCCTTTAGGCAGAGGACAGACCTTAAAAAACATCAGAAAATGCATAC
CGAAGAGAAAACCTTATGAATGTAATGAATGTGGGAAAGCCTTTAGCCAGAGCACATATCT
TACAAAACACCAAAAAATTCATAGTGAAGAGAAATCAAATATACATACTGAGTGTGGGGA
AACCATTAGACAAAACCTCTCTTTTTACAACAATAAAACCTCACACTGGAGAGTTCTCTG
25 AATGCCTTAAGAATTTGGTTAATATGGAGACCCTTCCCAGGGAAACAGAAGGAGGATCGT
GAAAACCGTTGACTACTTGAATGATCACATGGTTTAGTGGAGAGAGCATGATTCTGGGTT
TTAAAAGTCATGGATCTCAATCTCAGCTCCTATTACTAACTAGATCTTTTACTTTGGGGT
AAGTCACTTCATATCTTTAGGCCTTAATTTCTCATCTGAAAACCTGGAAGGCCTGACTTG
ACTTGTTGAGCTTAAGATCCTCAATTATTATATTTACTAGGAATTCAGTTTCTATAGAT
30 GTGGTTCAGAATTGTGACTTATTTATTGTACATCAGGTGTGATTCACAAGTGAGCTTGTA
GTAGTTATTAAGGAGTCAATAAAGATATGATATAAAAAAAAAAAAAAAAAA

AI688494 (IMAGE Clone ID: 2330499)

CATTTTCATCTTCATTGGATAGTGTTACATAGTAATATATTTATGTTTTCTTTAATCATT
35 TCATAACTTGGAATACTAACATAGTCAAACTCTAGGGTAGGTGATACATGAGTTTCT
GTAGTAATCTGGTTGGAGACATGTTGTAATTCTGTATATATATGTACATTTATCCCATGC
ATGTTATGCCTAAACTAAGACGGATACCCCTGAATTAAGAGGTGCTGTTATACATTGACC
AGGCTTAAGAATATCTCTTTAAAGTGTGTCGACATTTAATTGACCTTTGGAAGTTCATTC
TGTTAATCATACTCAAAGTGCTAAAGCTATGGTTGACTGCTCTGGTGTTTTTATATTCAT
40 TCGTGCTTTAGCATATAAATCTTCAGCATAATTGCTACTTATTTAGCAAGAGTTTCCTT
TATTTGAAAATGTGAGTTGTGCTTGTATTTTTGTGTCTTTCTTTCTTTCTTTTTTT
AACTTTGCTTCAGGCTGGGTAGTGGTAGAGGTTGAATTAAATGTTTTCTGTCAGTA
AAAAAAAAAAAA

AL157459

GAGCGAGCCCAGCAGCTTGCCCTTGACAGGTGGGGGCTGGCTGGGGCCTTAATGTGAAAA
GACAGTGGCAGGCAGCTGGAGTAGAGCGAGCCCAGCAGCCCTAAAAGGCTGCCTTCATGG
5 CCATCTAGCCCCAGTTCAGGGCAGCATCCATAGCCCACAAGCCAGCGTGGGTGGGGCGGG
GGTGGTCCCACAGCTGGGTTCACCTGAAGAGCCTCCGTGCCTCGGAGCAGGAGAGGCAG
GCTATGGCTGTCACCCTCCCTCCTGCCTGTGTCCAGTGAGAACTGACCTGAGTCCCCTT
CCAAACCCAGACCCACCTCCTGCCCCAGGCCCACTGAAGCATGTTCCATTTCTAAAAAGC
CCAGAGTTCAGTGTGTCCCAAGGAAAACCCAAAGTGGAGGTGCTCAGGTCCAGGGGAGTC
10 CAGTGGGCAGGACCCTTGGCAGGCAAGCCCCCTCCCTTCACTCCCAGGACCTACCTTCTGC
TAGTAAAGGACTGGCTTCATTCTAATTATGGCCCACAGACTGCCCCGGAGACCTGGAGGA
CAGCAGTGCTGGCACTTGGGTGTCCATGGGCCCCGTCTGCCGGCTCTGCCTGTGCTGCAAG
TGTTGGCCGTGGGTCCAGCCAACAACCTCCCTACGTCCTGTGTGGGGCCCTGCCCAAGTGG
ATGAGGCATTCCCTTGAGGAGTATCATTTTCCCTGACAATCCCCATCACCTTTAGGGGTTT
15 CCTGCTTGGCTCCTTTCCAGCTGAAAACTAGACCTGTGCCATTGGGGAAGCTGGACAAA
GTCTAGGGGGCCCGCCTGGTAGAGGGTCCCGGGAAGCTGGATCTGTCAGCCTCGGCCCTG
AGGCCCCCTGTAACTCAAGACTGTGAGCTGCCTCTAGGTGGTCACGTCTGGGAGCTAGCT
TGTATGGCTTCTGACCAGTATCAGGATTTCTGTTCTGAGAGCAGCGTGGGCAGCAAGGCA
GGGCAGCCCAGAGGTGGCAGCGGCAGGCAATCTGGTCACTAGGTCTTTGTGATGCCAAAA
20 ATAAAAGAGGGTGGGGTGGGTGCTTTCTGTTCTCTGATTGGATGGAGTCCGCCAGCAGG
CATGGGGCTACATTCCAGTGCCTGACTATAGGGAGGCACTCCTGATTCCATGGAGCAGCC
CGGACTTTGAGAATGGGCTCTGGTTTGCGGGGGGCAGGCGTACCAGACTGCAAGACCCCC
CAGTACCTCACCGTGCCAAATAGGAAGAGGTGGCCTTGGTGTAGCCAAATGGATCTTTTT
AACAGTGTGCCTTTGGGGAGGGACCCATGTCCATGGCTTCGTTGAGGGCCATCCATATGC
25 CAGCTGGGGGCCAGCCCACAGTGGCCATATTGGCTGCAGCAGGAATGGTGCCACCTCGG
CGAATTGAAGGGCTAAGAGTCCAGATAGCTAGGCCAGAGCTGGAAGCAGACAGTAAGGG
GAAGAGCTGCTCCCACAGGAGAGGGAGAGATTCCAGCTCACTGCGCAGCCTGGGAGGAGG
CGTGGATCCTGGCACGCTGAGCCTCAGGCACCAGCCTCCCTGTGCTCGACAGCAAAGTCT
TGACTCCTTCCTGCTGAGCACTGTGCTACCTTCACTGCTCCAAAGCCAGACTAACAGCTC
30 TCCAAGCCCTTGGGGTGA CTGCGCTTCCAGGAGCTGTTGGAGAAATGAGGATGTCTGTCC
CTGTCTGCCTGGGCAGGCCAGATTCCCTCCCCAGCAGCCGGGTCTCTCCAGACCCTGATTC
GGTGCCTTTCTGTTTACCAGCTACTTCAATCCCAAAGTTTGAATCTGCAGATACCTTACT
CCCAGCCACTTTGCCTTCTTACTGTGTTGTGTGTTTTTCTGGTGCTTCAAGAGCGTGTG
CAGGGCAAGTGCCGTCAGTGGGAACTGCACCAGATGCTCAGACTTGGTTGTCTTATGTTT
35 ACCAATAAAATAAAAGTAGACTTTTTCTATTTTTATTTGCTGCTATTTGTGTGTGTGTTG
TGTTTGTGTAGCTAGGTATCTGGCACTTCTGACGATGCATTGTTGCTTTTTTCCCGAAGG
TCCCGCAGGAACTGTGGCAATGGTGTGTGTGTGAAATGGTGTGTTAACCGCGTTTTGTTT
GCTCCTGTATTGAATAGGAAGCAGTGGCCAGTCTGTCTTCCTTAGAGATGTTAGCATATT
TTTATATGTATATATTTTTGTACCAAAAAAGAGTGTTCCTTGTTTTGGTTACACTCGAAAT
40 TCTGACCTAGCTGGAGAGGGCTCTGGGCCGAGAGCTTTCATAAGGGGAGACTTCAGGGG
AGGATCAAGCTTTGAACCAAAGCCAATCACTGGCTTGATTTGTGTTTTTTAATTAAAAAA
AAAAATCATTCATGTATGCCACTTCTAAAAA

BC002480 (IMAGE Clone ID: 3350037)

GGCACGAGGCTGAGACCGGTGCGCCGCGCGCTAGTGGCCGCTCTTCCGCGGGCTAGCGGG
CGGTGGGGGCGCCAGCAGCGCGGAAGGCGGGCACGCGGGCCATGGCTCCCTGGGCGGAGG
CCGAGCACTCGGCGCTGAACCCGCTGCGCGCGGTGTGGCTCACGCTGACCGCCGCCTTCC
5 TGCTGACCCTACTGCTGCAGCTCCTGCCGCCCGGCCTGCTCCCGGGCTGCGCGATCTTCC
AGGACCTGATCCGCTATGGGAAAACCAAGTGTGGGGAGCCGTCGCGCCCCGCGCCTGCC
GAGCCTTTGATGTCCCCAAGAGATATTTTTCCCACTTTTATATCATCTCAGTGCTGTGGA
ATGGCTTCCTGCTTTGGTGCCTTACTCAATCTCTGTTCTTGGGAGCACCTTTTCCAAGCT
GGCTTCATGGTTTGCTCAGAAATCTCGGGGCGGCACAGTTCCAGGGAGGGGAGCTGGCAC
10 TGTCTGCATTCTTAGTGCTAGTATTTCTGTGGCTGCACAGCTTACGAAGACTCTTCGAGT
GCCTCTACGTGAGTGTCTTCTCCAATGTCATGATTCACGTCGTGCAGTACTGTTTTGGAC
TTGTCTATTATGTCCTTGTGGCCTAACTGTGCTGAGCCAAGTGCCAATGGATGGCAGGA
ATGCCTACATAACAGGGAAAAATCTATTGATGCAAGCACGGTGGTTCCATATTCTTGGGA
TGATGATGTTTCATCTGGTCATCTGCCCATCAGTATAAGTGCCATGTTATTCTCGGCAATC
15 TCAGGAAAAATAAAGCAGGAGTGGTCATTCACTGTAACCACAGGATCCCATTTGGAGACT
GGTTTGAATATGTTTCTTCCCCTAACTACTTAGCAGAGCTGATGATCTACGTTTCCATGG
CCGTCACCTTTGGGTTCACAACTTAACTGGTGGCTAGTGGTGACAAATGTCTTCTTTA
ATCAGGCCCTGTCTGCCTTTCTCAGCCACCAATTCTACAAAAGCAAATTTGTCTCTTACC
CGAAGCATAGGAAAGCTTTCCTACCATTTTTGTTTTAAGTTAACCTCAGTCATGAAGAAT
20 GCAAACCAGGTGATGGTTTCAATGCCTAAGGACAGTGAAGTCTGGAGCCCAAAGTACAGT
TTCAGCAAAGCTGTTTGAACTCTCCATTCCATTTCTATACCCCAAGTTTTCACTGAA
TGAGCATGGCAGTGCCACTCAAGAAAATGAATCTCCAAGTATCTTCAAAGAATAAATAC
TAATGGCAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA